Strategic Experimentation

Bolton, P.; Harris, C.

Publication date:
1996

1.4

Strategic Experimentation: a Revision*

Patrick Bolton
CentER for Economic Research, Tilburg University
and ECARE, Université Libre de Bruxelles

Christopher Harris
King's College, Cambridge

December 1992
Revised February 1993
Revised December 1994
Revised November 1995

Abstract

This paper extends the classic two-armed bandit problem to a many-agent setting in which I players each face the same experimentation problem. The main change from the single-agent problem is that an agent can now learn from the current experimentation of other agents. Information is therefore a public good, and a free-rider problem in experimentation naturally arises. More interestingly, the prospect of future experimentation by others encourages agents to increase current experimentation, in order to bring forward the

*We would like to thank the Studienzentrum, Gerzensee, where a substantial part of the work reported in this paper was undertaken. We would also like to thank David Cox, Drew Fudenberg, Graham Hill, Meg Meyer, Khalid Sekkat, Neil Shephard, Hyun Shin, Richard Spady, Jean Tirole and especially Douglas Gale and three anonymous referees for their help, and seminar participants at Berkeley, Cambridge, Chicago, ECARE, LSE, Minnesota, MIT, Northwestern, Oxford, Paris, Princeton, Stanford, Toulouse, UCLA, University College London and Yale for their comments.

time at which the extra information generated by such experimentation becomes available. The paper provides an analysis of the set of stationary Markov equilibria in terms of the free-rider effect and the encouragement effect.

The paper is a revision of our earlier paper, Bolton and Harris [7]. The main modification concerns the formulation of randomization in continuous time. C.f. Harris [12]. The earlier paper explored one formulation based on the idea of rapid alternation over the state space. The current paper explores a formulation which is the closest analogue of the discrete-time formulation. It is based on the idea of randomization at each instant of time.

1 Introduction

This paper analyses a game of strategic experimentation in which individual players can learn from the experiments of others as well as their own. Given that experimentation typically entails an opportunity cost, and that information obtained from an experiment is valuable to all players, individual players attempt to free ride on the experiments of others. This informational externality drives a wedge between equilibrium experimentation and socially optimal experimentation. On the other hand, an individual player may be encouraged to experiment more if, by so doing, she can bring forward the time at which the information generated by the experimentation of others becomes available. This encouragement effect mitigates the free-rider effect. The objective of the paper is to analyze socially optimal and equilibrium experimentation strategies in terms of the free-rider and the encouragement effects.

The game of strategic experimentation we consider is a many-player common-value extension of the classic continuous-time two-armed bandit problem as presented in Karatzas [14] and Berry and Fristedt [4]. At any given time, each player in our game must choose between the 'safe' action and the 'risky' action. The underlying payoff of the safe action is known and common to all players. The underlying payoff of the risky action is unknown but common to all players, and it can be either higher or lower than that of the safe action.

The actual payoff obtained by a player from an action is the underlying payoff of that action plus noise. Once the actions have been chosen and the payoffs realized, all players observe all actions chosen and all payoffs. They therefore obtain information about the underlying payoff of the risky action by observing the payoffs obtained by those players who chose to experiment. They also observe a background signal. This signal provides them with further information about the underlying payoff of the risky action. In particular, it ensures that some information is obtained even when no player chooses to experiment.

In the case of socially optimal experimentation, we restrict attention to symmetric optima. The socially optimal experimentation policy is very simple. There is a cutoff $C_* \in (0,1)$ such that no player experiments when the common belief p that the unknown payoff is greater than the known payoff falls below C_*, and such that all the players experiment when p exceeds C_*.

Simple and intuitive comparative statics results about the socially optimal solution are also obtained, the most interesting being that the payoff of a representative player is increasing in both the number of players and the quality of background information.

These results are easy to explain. The payoff of a representative player is increasing with the number of players because of the public good nature of the information produced through experimentation: the larger the number of players, the smaller the contribution of any one player to any given level of provision of the public good. The result that the payoff of a representative player is increasing in the quality of background information is a direct consequence of the envelope theorem: the shadow value of information is positive and therefore more background information increases the payoff.

In the case of equilibrium experimentation, we restrict attention to symmetric equilibria in stationary Markov strategies. Associated with any symmetric equilibrium there are cutoffs C_1 and C_N. Players experiment with probability 0 when p falls below C_1, with probability rising continuously from 0 to 1 as p rises from C_1 to C_N, and with probability 1 when p exceeds C_N. Because of free riding, the payoff of a representative player in a symmetric equilibrium is strictly less than the full-information payoff, even in the limit as the number of players tends to infinity. This is a stark illustration of the strength of the free-rider effect. However, despite free riding, the symmetric-equilibrium payoff is rising both in the number of players and in the quality of background information.

The latter results are by no means obvious, since total information is not monotonic either in the number of players or in the quality of background information. Indeed, when there is no discounting, total information is independent of the number of players for low values of p, decreasing for intermediate values of p, and increasing for high values of p. Similarly, again when there is no discounting, total information is increasing in the quality of background information for low values of p, decreasing for intermediate values of p, and increasing for high values of p.

They can, however, be explained by decomposing the different effects of an increase in the number of players or in background information on the payoff of a representative player into three parts: the direct effect of the change on her objective, the indirect effect on her objective of the change in the behavior of the other players, and the indirect effect on her objective of the change in her own behavior. The second effect is a strategic effect. The

third is always zero by the envelope theorem.

Increasing the number of players has no direct effect on the objective of a typical player. As for the strategic effect, total experimentation by the other players is unchanged at 0 for low values of p, and increases from $N-1$ to N for high values of p. For intermediate values of p, there is both a free-rider and an encouragement effect. However, from the point of view of any given player, the reduction in the per capita experimentation of the other players that results from free riding on the new player is exactly offset by the increase in the number of players. The encouragement effect therefore dominates. Overall, then, experimentation by the other players rises.

Increasing background information has a direct positive effect on the objective of a typical player. It does, however, have a potentially ambiguous effect on the experimentation by the other players. They will want to free ride on the extra information available currently, but they will also want to experiment more, to take better advantage of the extra information available in the future. From the point of view of the chosen player, however, the reduction in the experimentation of the other players that results from free riding is exactly offset by the increase in background information. It is therefore the encouragement effect that wins out overall.

Two aspects of our modelling strategy deserve comment: our use of a continuous-time formulation of the two-armed bandit problem, and our introduction of background information.

We work with a continuous-time formulation of the two-armed bandit problem for reasons of tractability. This formulation allows us to obtain closed-form solutions for the socially efficient experimentation strategy and for the associated value function. Similarly, in the special case with no discounting, it allows us to obtain closed-form solutions for equilibrium experimentation strategies and the associated value functions. Such a sharp characterization could not be obtained in a discrete-time formulation.

We work with background information for two main reasons. First, in the presence of background information, the undiscounted case is non-degenerate. This allows us to begin our analysis of strategic experimentation with consideration of an easier special case. Secondly, much of the literature on optimal experimentation has focused on the question of whether complete learning takes place in the long run. See Rothschild [20], McLennan [17], Easley and Kiefer [9] and Aghion, Bolton, Harris and Jullien [1], for example. In the presence of background information, this question does not arise: no matter

what experimentation strategy players follow, they end up acquiring complete information. This leads us to focus on a different question, namely how fast players can and ought to learn. Answering this question amounts to determining equilibrium and socially optimal experimentation strategies. Interestingly enough, it turns out that the socially optimal experimentation strategy is qualitatively very similar to that for the case of no background information.

The literature on social learning or learning by experimentation in many-player settings is growing rapidly. To our knowledge only one other paper (Smith [22]) considers a similar framework to ours. Smith focuses on limit beliefs and does not attempt to characterize socially optimal or equilibrium experimentation strategies. Several other papers on social learning in settings with a pure informational externality deal with issues related to ours. Hendricks and Kovenock [13] consider a finitely repeated game of oil exploration between two oil companies owning adjacent oil fields. Ellison and Fudenberg [10] analyze an infinitely repeated game of social learning in which individual players can learn from their own experience and that of their neighbors. Players follow intuitively plausible rules of thumb to determine their experimentation actions. Chamley and Gale [8] emphasize the free riding aspect of social learning in a game of timing of investments.

A second set of papers on social learning combines the informational externality of experimentation with other strategic interactions among players. Thus, Bhattacharya, Chatterjee and Samuelson [5] consider a game of strategic research and development. Mirman, Samuelson and Urbano [18] consider a game of duopoly signal jamming in which the experiments of one player partly serve the purpose of confusing the other player. Rob [19] studies a game of entry with unknown market size. In this game, each entrant imposes two externalities on the other players, a positive informational externality on future potential entrants who learn something about the size of the market following entry, and a negative externality on incumbents who see their profits reduced as a result of entry. Finally, Aghion, Espinosa and Jullien [2] consider a Bertrand pricing game in which equilibrium price dispersion is the result of learning by experimentation. As these models combine different strategic interactions besides the informational externality it is difficult to compare them with one another.

A final set of papers to which our work is related is the recent literature on herd behavior. These papers consider social learning when the payoffs

associated with actions are not perfectly observable by all players. Banerjee [3] and Bikhchandani, Hirshleifer and Welch [6] consider a one-shot game in which players move sequentially and observe the moves of the players ahead of them but not their payoffs. Both papers show that inefficient herd behavior may occur in this setting. These papers raise the question of the relation between free-riding and herding. Also, an open question is to what extent herding phenomena survive in a repeated setting. This question has recently been addressed by Vives in a setting without experimentation [23].

The paper is organized as follows. Section 2 describes the model. Section 3 solves the filtering problem for any profile of experimentation actions. Section 4 sets up the optimization problem faced by the players when they act as a team. The team problem can be described by a Bellman equation. Section 5 uses an alternative form for the Bellman equation to characterize the team solution. Section 6 gives the comparative statics of the team solution. Section 7 illustrates and discusses the two main informational externalities between players in this model: the free-rider and encouragement effects. Section 8 briefly analyses strategic experimentation in the special case with two players and no discounting. Section 9 discusses mixed strategies. Section 10 characterizes symmetric equilibria in the general case with many players and discounting. Section 11 gives the comparative statics of the strategic problem. Section 12 shows that our mixed-strategy equilibria can also be interpreted as public-randomization equilibria. Section 13 explains how our model can be generalized to the case where the payoff of the unknown action can take on many values (rather than just two), and where players may observe more than one signal.

2 The Model

There are N identical infinitely lived risk-neutral players. At time t, each player i chooses between two actions:

- a 'safe' action, with payoff $d\pi_i(t) = s\,dt + \sigma\,dZ_i(t)$, and
- a 'risky' action, with payoff $d\pi_i(t) = \mu\,dt + \sigma\,dZ_i(t)$.

These choices are made simultaneously and independently. All players then observe all actions chosen and the resulting payoffs. They also observe

- a background signal $d\pi_0(t) = \sqrt{\xi}\mu\, dt + \sigma\, dZ_0(t)$.

Here:

- s is fixed and known;
- $\mu \in \{\ell, h\}$ is unknown;
- $\ell < s < h$;
- $\xi > 0$ is the quality of the background signal;
- the $dZ_i(t)$ are independently and normally distributed with mean 0 and variance dt for $0 \leq i \leq N$.

Player i's objective is to maximize the expectation of the present discounted value of his payoff stream, namely

$$\mathrm{E}\left[\int_0^\infty re^{-rt} d\pi_i(t)\right],$$

where $r > 0$ is the discount rate.

In this model player i observes the background signal $d\pi_0(t)$ and the payoffs $d\pi_j(t)$ of all players j including herself. The background signal $d\pi_0(t)$ is composed of the deterministic contribution $\sqrt{\xi}\mu\, dt$ and the stochastic shock $\sigma\, dZ_0(t)$. The first contribution ensures that it conveys some information about μ. The second ensures that this information is noisy. The payoff $d\pi_j(t)$ of player j conveys no information in the case where player j plays safe. In the case where player j plays risky, $d\pi_j(t)$ conveys noisy information about μ. Since the background signal is always observed, player i always obtains some information. Hence she will eventually learn the value of μ more or less exactly, and incomplete learning cannot occur in this model. The problem is, rather, to determine how quickly players learn the value of μ.

3 The Filtering Problem

In order to characterize the optimal experimentation policy of a single player or of the team of N players it is helpful to determine how the players' common belief $p(t)$ that μ is high evolves as more information about μ becomes

available. One of the advantages of our continuous-time formulation is that it is possible to give a simple answer to this question.

Let 0 denote the safe action, let 1 denote the risky action, and let $x_i \in \{0,1\}$ denote the action taken by player i. In other words, let x_i denote the amount of experimentation undertaken by player i. Let $p(t)$ denote the prior belief that μ is high at the outset of instant t, let $p(t+dt)$ denote the posterior belief that μ is high at the conclusion of instant t, and let

$$dp(t) = p(t+dt) - p(t)$$

denote the change in beliefs concerning μ. Finally, let

$$\Sigma(p) = \left(p(1-p)\left(\frac{h-\ell}{\sigma}\right)\right)^2.$$

Then we have the following lemma.

Lemma 1 $dp(t) \sim \mathrm{N}\left[0, \left(\xi + \sum_{i=1}^{N} x_i\right)\Sigma(p(t))\,dt\right]$.

Note first that beliefs can be expected to follow a martingale. In other words, the expectation of $p(t+dt)$ conditional on current information should be $p(t)$. Or, equivalently, the expectation of $dp(t)$ conditional on current information should be 0. Lemma 1 confirms that this is indeed the case. Secondly, the better the information received about μ, the higher the variance of the posterior should be. In particular, the variance of the posterior should be higher: the higher the quality ξ of the background signal; the larger the number $\sum_{i=1}^{N} x_i$ of players experimenting; and the higher the signal-to-noise ratio $\frac{h-\ell}{\sigma}$. Lemma 1 confirms these intuitions too. Finally, Lemma 1 makes clear that the posterior is unchanged from the prior when there is already certainty as to which state of the world obtains, i.e. whenever $p \in \{0,1\}$.

Proof. Players observe the background signal

$$d\pi_0(t) = \sqrt{\xi}\mu\,dt + \sigma\,dZ_0(t),$$

and the payoffs

$$d\pi_i(t) = ((1-x_i)s + x_i\mu)\,dt + \sigma\,dZ_i(t)$$

for $1 \leq i \leq N$. These signals are observationally equivalent to the signals

$$d\widetilde{\pi}_i(t) = x_i\widetilde{\mu}\,dt + dZ_i(t)$$

for $0 \leq i \leq N$, where $x_0 = \sqrt{\xi}$ and $\widetilde{\mu} = \frac{\mu - s}{\sigma}$.

Now $\widetilde{\mu}$ takes the values $\widetilde{\ell} = \frac{\ell - s}{\sigma}$ and $\widetilde{h} = \frac{h - s}{\sigma}$ with probabilities $(1-p)$ and p, and the $dZ_i(t)$ are independently and normally distributed with mean 0 and variance dt. Hence, applying Bayes' Rule, we obtain

$$p(t+dt) = \frac{p(t) F\left(\widetilde{h}\right)}{p(t) F\left(\widetilde{h}\right) + (1 - p(t)) F\left(\widetilde{\ell}\right)},$$

where

$$F(\widetilde{\mu}) = (2\pi\, dt)^{-N/2} \exp\left(-\frac{1}{2\, dt} \sum_{i=0}^{N} \left(d\widetilde{\pi}_i(t) - x_i \widetilde{\mu}\, dt\right)^2\right)$$

is the probability of observing the payoff profile $d\widetilde{\pi}(t) = \times_{i=0}^{N} d\widetilde{\pi}_i(t)$ given $\widetilde{\mu}$. Hence

$$dp = \frac{p(1-p)\left(\widetilde{F}\left(\widetilde{h}\right) - \widetilde{F}\left(\widetilde{\ell}\right)\right)}{p\widetilde{F}\left(\widetilde{h}\right) + (1-p)\widetilde{F}\left(\widetilde{\ell}\right)},$$

where

$$\widetilde{F}(\widetilde{\mu}) = \exp\left(\sum_{i=0}^{N} x_i \widetilde{\mu}\, d\widetilde{\pi}_i - \frac{1}{2} \sum_{i=0}^{N} x_i^2 \widetilde{\mu}^2\, dt\right),$$

and where we have suppressed dependence on t.

Next,

$$\widetilde{F}(\widetilde{\mu}) = 1 + \sum_{i=0}^{N} x_i \widetilde{\mu}\, d\widetilde{\pi}_i - \frac{1}{2} \sum_{i=0}^{N} x_i^2 \widetilde{\mu}^2\, dt + \frac{1}{2}\left(\sum_{i=0}^{N} x_i \widetilde{\mu}\, d\widetilde{\pi}_i\right)^2$$

(neglecting terms of order $dt^{\frac{3}{2}}$ and higher)

$$= 1 + \sum_{i=0}^{N} x_i \widetilde{\mu}\, d\widetilde{\pi}_i$$

(since $d\widetilde{\pi}_i d\widetilde{\pi}_j = dt$ if $i = j$ and $d\widetilde{\pi}_i d\widetilde{\pi}_j = 0$ if $i \neq j$). Hence

$$dp = \frac{p(1-p)\left(\widetilde{h} - \widetilde{\ell}\right) \sum_{i=0}^{N} x_i\, d\widetilde{\pi}_i}{1 + \sum_{i=0}^{N} x_i \widetilde{w}(p)\, d\widetilde{\pi}_i}$$

(where $\widetilde{w}(p) = (1-p)\widetilde{\ell} + p\widetilde{h}$)

$$= p(1-p)\left(\widetilde{h} - \widetilde{\ell}\right)\left(\sum_{i=0}^{N} x_i d\widetilde{\pi}_i\right)\left(1 - \sum_{i=0}^{N} x_i \widetilde{w}(p)\, d\widetilde{\pi}_i\right)$$

(neglecting terms of order $dt^{\frac{3}{2}}$)

$$= p(1-p)\left(\widetilde{h} - \widetilde{\ell}\right)\left(\sum_{i=0}^{N} x_i d\widetilde{\pi}_i - \sum_{i=0}^{N} x_i^2 \widetilde{w}(p)\, dt\right)$$

(noting once again that $d\widetilde{\pi}_i d\widetilde{\pi}_j = dt$ if $i = j$ and $d\widetilde{\pi}_i d\widetilde{\pi}_j = 0$ if $i \neq j$)

$$= p(1-p)\left(\widetilde{h} - \widetilde{\ell}\right)\sum_{i=0}^{N} x_i\, d\widetilde{Z}_i$$

(where $d\widetilde{Z}_i = d\widetilde{\pi}_i - x_i \widetilde{w}(p)\, dt$).

Finally, the expectation of $d\widetilde{Z}_i$ conditional on the information available to the players at time t is 0, and $d\widetilde{Z}_i d\widetilde{Z}_j = dt$ if $i = j$ and $d\widetilde{Z}_i d\widetilde{Z}_j = 0$ if $i \neq j$. That is, the profile $\widetilde{Z} = \times_{i=0}^{N} \widetilde{Z}_i$ follows a standard $(N+1)$-dimensional Wiener process relative to the players' information. Hence dp has mean 0 and variance

$$\left(p(1-p)\left(\widetilde{h} - \widetilde{\ell}\right)\right)^2 \left(\sum_{i=0}^{N} x_i^2\right) dt.$$

Recalling that $x_0 = \sqrt{\xi}$, that $x_i \in \{0,1\}$, that $\widetilde{\ell} = \frac{\ell - s}{\sigma}$ and that $\widetilde{h} = \frac{h-s}{\sigma}$, we obtain the required conclusion. ■

4 The Team Problem

Suppose that the players act as a team. Then the problem reduces to that of choosing the number $n \in \{0, 1, ..., N\}$ of experiments to maximize the average payoff. If n experiments are chosen, then the change in beliefs

$$dp \sim \mathrm{N}\left[0, (\xi + n)\,\Sigma(p)\, dt\right].$$

Moreover the expectation, given current information, of the flow payoff per capita is

$$\frac{N-n}{N}\, s + \frac{n}{N}\, w(p),$$

where $w(p) = (1-p)\,\ell + p\,h$ is the expectation, given current information, of the flow payoff from the risky arm. The optimal experimentation problem therefore reduces to the problem of controlling the variance of the diffusion process p. The latter problem is amenable to the techniques of dynamic programming.

Lemma 2 *The value function $u_* : [0,1] \to R$ for the team problem is the unique bounded solution of the Bellman equation*

$$u_* = \max_{n \in \{0,1,\ldots,N\}} \left(\left(\frac{N-n}{N}\, s + \frac{n}{N}\, w \right) + \frac{1}{r}\,(\xi + n)\,\Sigma \frac{u_*''}{2} \right), \tag{1}$$

where we have suppressed the dependence of Σ, u and w on p.

In particular,

$$u_*(0) = \max_{n \in \{0,1,\ldots,N\}} \left(\frac{N-n}{N}\, s + \frac{n}{N}\, \ell \right) = s$$

and

$$u_*(1) = \max_{n \in \{0,1,\ldots,N\}} \left(\frac{N-n}{N}\, s + \frac{n}{N}\, h \right) = h,$$

since $\Sigma(0) = \Sigma(1) = 0$.

Proof. Suppose that the current belief is $p \in [0,1]$, that n members of the team experiment, and that continuation payoffs are given by $c : [0,1] \to R$. Then the payoff from the current instant of time dt is

$$r \left(\frac{N-n}{N}\, s + \frac{n}{N}\, w(p) \right) dt, \tag{2}$$

and the payoff from the remainder of time is

$$e^{-r\,dt}\, c(p + dp).$$

Now

$$e^{-r\,dt} = 1 - r\,dt$$

and

$$c(p + dp) = c(p) + c'(p)\,dp + \frac{1}{2} c''(p)(dp)^2.$$

Moreover dp is normally distributed with mean 0 and variance $(\xi + n)\,\Sigma\,(p)\,dt$. Hence $E\,[dp] = 0$, $E\left[(dp)^2\right] = (\xi + n)\,\Sigma\,(p)\,dt$, and the expectation of the continuation payoff is

$$(1 - r\,dt)\left(c(p) + \frac{1}{2}c''(p)\ (\xi + n)\,\Sigma\,(p)\,dt\right). \tag{3}$$

Adding (2) and (3) we obtain the expectation of the overall payoff, namely

$$H\,(n, p, c\,(\cdot)) = c(p) + r\,dt\,\left(\left(\frac{N-n}{N}\,s + \frac{n}{N}\,w\,(p)\right) + \frac{1}{r}\,(\xi + n)\,\Sigma\,(p)\,\frac{c''(p)}{2} - c(p)\right),$$

where we have dropped terms of order $(dt)^2$. Finally, the value function u_* for the team problem is, as usual, the unique bounded solution of the Bellman equation

$$u_*\,(p) = \max_{n \in \{0,1,\ldots,N\}} H\,(n, p, u_*\,(\cdot))\ \text{ for all } p \in [0,1]\,.$$

It is easy to see that this equation reduces to (1). ■

In a discrete-time setting, the Bellman equation states that the current payoff is equal to the maximum over the control variable of the expectation of the current payoff plus the discounted value of the continuation payoff. In the present, continuous-time, setting:

- u_* is the current payoff;
- n is the control variable;
- $\dfrac{N-n}{N}\,s + \dfrac{n}{N}\,w$ is the expectation of the current flow payoff;
- $\dfrac{1}{r}$ is the discount factor;
- $(\xi + n)\ \Sigma\dfrac{u_*''}{2}$ is the expectation of the rate of change of the continuation payoff.

The Bellman equation therefore states that the current payoff is the maximum over the control variable of the expectation of the current flow payoff plus the discounted value of the *rate of change* of the continuation payoff.

The Bellman equation also tells us how to choose the optimal policy n_*. We should set

$$n_* = \left\{ \begin{array}{lll} N & \text{if} & \frac{1}{r}\Sigma\frac{u_*''}{2} > \frac{s-w}{N} \\ 0 & \text{if} & \frac{1}{r}\Sigma\frac{u_*''}{2} < \frac{s-w}{N} \end{array} \right\}.$$

Notice that:

- $\frac{1}{r}$ is the discount rate;
- Σ is the amount of information revealed by any given experiment;
- $\frac{u_*''}{2}$ is the shadow value of information;
- $s - w$ is the opportunity cost of experimentation.

We may therefore interpret

$$\frac{1}{r}\Sigma\frac{u_*''}{2}$$

as the shadow value of experimentation, and the Bellman equation tells us that we should maximize experimentation if the shadow value of experimentation exceeds its opportunity cost, and minimize experimentation otherwise.

Remark 1 *One might have expected that experimentation would take place if and only if the posterior is sufficiently high. This is not immediately apparent from the present formulation of the Bellman equation. It is, however, immediate from the alternative formulation introduced below.*

Remark 2 *One might also have expected to see levels of experimentation intermediate between* 0 *and* N*. However, in the continuous-time setting considered here, the marginal value of an additional experiment is constant. This seems to be because the time interval* dt *during which the team experiments is infinitesimal.*

5 Alternative Formulation of the Bellman Equation

There are several different ways of formulating the Bellman equation for the team problem. In this section we introduce an alternative formulation to the one given above, which yields further insight into the team problem.

Lemma 3 *The value function $u_* : [0,1] \to R$ for the team problem is the unique bounded solution of the Bellman equation*

$$0 = \max_{n \in \{0,1,\ldots,N\}} \left(-\frac{r\,(s-w)}{N} + \frac{r}{\xi+n}\left(\frac{\xi\,(s-w)}{N} - (u_* - s)\right) + \Sigma\frac{(u_* - s)''}{2}\right). \tag{4}$$

Lemma 3 tells us that the optimal-experimentation problem is equivalent to a randomized-stopping problem. In this problem:

- all payoffs are measured relative to the safe payoff s;
- there is no discounting;
- the state variable p evolves according to the equation $dp \sim \mathrm{N}\left[0, \Sigma\,(p)\,dt\right]$ until such time as the process is stopped;
- the flow payoff $-\dfrac{r\,(s-w)}{N}$ is obtained up to the point at which the process is stopped;
- the process is stopped with probability $\dfrac{r}{\xi+n}$ per unit time;
- when the process is stopped, a lump-sum payoff of $\dfrac{\xi\,(s-w)}{N}$ is obtained.

In other words, r times the social opportunity cost of experimentation $\frac{s-w}{N}$ is incurred up to the time at which the process is stopped. A lump-sum benefit equal to the value of the background information ξ in terms of the social opportunity cost of experimentation is then obtained.

Lemma 3 also tells us that the optimal policy for this randomized-stopping problem is given by:

$$n_* = \left\{ \begin{array}{lll} N & \text{if} & u_* - s > \dfrac{\xi}{N}(s-w) \\ & & \\ 0 & \text{if} & u_* - s < \dfrac{\xi}{N}(s-w) \end{array} \right\}. \tag{5}$$

Since $u_* - s$ increases strictly from 0 to $h - s$ as p increases from 0 to 1, and $s - w$ decreases strictly from $s - \ell$ to $s - h$ as p increases from 0 to 1, this implies that there is a unique cutoff $C_* \in (0, b)$ such that $n_* = 0$ if $p < C_*$ and $n_* = N$ if $p > C_*$, where b is the break-even point at which $s - w = 0$. (C.f. Lemma 4 below.)

Proof. The Bellman equation given in Lemma 2 can be written equivalently as

$$\max_n \left(\left(\frac{N-n}{N} s + \frac{n}{N} w \right) + \frac{1}{r}(\xi + n) \Sigma \frac{u_*''}{2} - u_* \right) = 0.$$

Rearranging slightly, we obtain

$$\max_n \left(-\frac{(\xi+n)(s-w)}{N} + \frac{\xi(s-w)}{N} - (u_* - s) + \frac{1}{r}(\xi+n)\Sigma\frac{u_*''}{2} \right) = 0.$$

Now, if the maximum of a function is zero, then the maximum and the set of maximizers is unchanged if the function is divided through by any other function, provided only that the second function is strictly positive. We may therefore divide the maximand in the latter equation by $\frac{\xi+n}{r}$. Doing so yields (4). ■

Remark 3 *In a discrete-time setting, we would expect to find a sequence of cutoffs C_{*1}, C_{*2}, ..., C_{*N} such that n players experiment iff $C_{*n} < p < C_{*(n+1)}$. Our main reason for expecting this is that, in a discrete-time setting, the marginal information from an additional experiment is decreasing in n (the number of players experimenting), while the marginal cost of experimentation is independent of n. We would therefore expect to see an interior solution for the number of players who play the risky action, at least for some values of p.*

Remark 4 *One can solve explicitly for u_* and C_*. Let*

$$\gamma_1 = \sqrt{1 + \frac{8r\sigma^2}{\xi(h-\ell)^2}}, \quad \gamma_2 = \sqrt{1 + \frac{8r\sigma^2}{(\xi+N)(h-\ell)^2}},$$

$$u_1(p) = p^{(\gamma_1+1)/2}(1-p)^{-(\gamma_1-1)/2} \quad \textit{and} \quad u_2(p) = p^{-(\gamma_2-1)/2}(1-p)^{(\gamma_2+1)/2}.$$

Then

$$C_* = \frac{(-N + \xi\gamma_1 + (\xi+N)\gamma_2)(s-\ell)}{(-N + \xi\gamma_1 + (\xi+N)\gamma_2)(s-\ell) + (N + \xi\gamma_1 + (\xi+N)\gamma_2)(h-s)}$$

and

$$u_*(p) = \left\{ \begin{array}{ll} s + \dfrac{\xi(s - w(C_*))}{N u_1(C_*)} u_1(p) & \textit{for} \quad p \in [0, C_*] \\ w(p) + \dfrac{(\xi+N)(s - w(C_*))}{N u_2(C_*)} u_2(p) & \textit{for} \quad p \in [C_*, 1] \end{array} \right\}.$$

This completely characterizes the team solution.

6 Comparative Statics of the Team Problem

In this section we briefly describe how the team value function and the team cutoff change with the discount rate, the number of agents in the team and the quality of background information.

Let b denote the break-even point at which $s - w = 0$, let

$$u_\infty(p) = \max\{s, w(p)\}$$

denote the myopic payoff, and let

$$u_0(p) = (1-p)\ s + p\,h$$

denote the full-information payoff. Then we have the following simple lemma.

Lemma 4 *For all $p \in (0,1)$:*

(i) $u_\infty(p) < u_*(p) < u_0(p)$;

(ii) $u_*''(p) > 0$.

In other words: the possibility of learning about the payoff of the risky arm allows the team to obtain a payoff strictly greater than the myopic payoff; the impossibility of learning the payoff of the risky arm instantaneously prevents the team from attaining the full-information payoff; and the value of information, as measured by the second derivative of the value function, is strictly positive as long as there is some uncertainty concerning the payoff of the risky arm.

Proof. The weak versions of the inequalities in the statement of the lemma are all easy consequences of the optimal-experimentation formulation of our problem. We begin by sketching these arguments. Note first that one valid strategy is always to play safe. This yields the payoff s. Another valid strategy is always to play risky. This yields the payoff w. Hence $u_* \geq \max\{s, w\} = u_\infty$. Secondly, any strategy that can be played when there is incomplete information can also be played when there is complete information. Hence u_* must be less than or equal to the full-information payoff u_0. Finally, the payoff from any given strategy is linear in the prior p. Hence the value function, which is the upper envelope of such payoffs, must be convex in p.

Let us turn now to the strict inequalities. Consider the myopic strategy of putting $n = 0$ if $p < b$ and $n = N$ if $p \geq b$. The value u of this strategy is the unique bounded solution of the equation

$$u = \left\{ \begin{array}{cc} s + \dfrac{1}{r}\xi\Sigma\dfrac{u''}{2} & \text{if} \quad p < b \\ \\ w + \dfrac{1}{r}(\xi + N)\Sigma\dfrac{u''}{2} & \text{if} \quad p \geq b \end{array} \right\}.$$

It is easy to check that $u \geq u_\infty$, with strict inequality on $(0,1)$. Since $u_* \geq u$, it follows that $u_* \geq u_\infty$, with strict inequality on $(0,1)$. Next, the randomized-stopping formulation of the Bellman equation (4) implies that

$$\frac{1}{r}\Sigma\frac{u_*''}{2} = \min\left\{\frac{u_* - s}{\xi}, \frac{u_* - w}{\xi + N}\right\}.$$

Since $u_* \geq u_\infty$, with strict inequality on $(0,1)$, it follows that $u_*'' \geq 0$, with strict inequality on $(0,1)$. Thirdly, the flow payoff associated with any strategy is at most u_0. Since u_0 is linear, it follows that the overall payoff u

associated with any strategy is also at most u_0. Hence $u_* \leq u_0$. Finally, suppose for a contradiction that $u_*(p) = u_0(p)$ for some $p \in (0,1)$. Then $u_* - u_0$ is maximized at p, and therefore $0 \geq (u_* - u_0)''(p) = u_*''(p)$. This is the required contradiction. ■

Using Lemma 4, one can establish the following comparative statics results for the equilibrium payoff.

Theorem 5 *For all $p \in (0,1)$:*

(i) *$u_*(p)$ is strictly decreasing in r, $u_*(p) \to u_0(p)$ as $r \to 0$, and $u_*(p) \to u_\infty(p)$ as $r \to \infty$;*

(ii) *$u_*(p)$ is strictly increasing in ξ, and $u_*(p) \to u_0(p)$ as $\xi \to \infty$;*

(iii) *$u_*(p)$ is strictly increasing in N, and $u_*(p) \to u_0(p)$ as $N \to \infty$.* ■

One can also establish the following results for the optimal cutoff.

Theorem 6 *Let C_* denote the optimal cutoff. Then:*

(i) *C_* is strictly increasing in r,*

$$C_* \to \frac{\xi(s-\ell)}{\xi(s-\ell) + (\xi+N)(h-s)}$$

as $r \to 0$, and $C_* \to b$ as $r \to \infty$;

(ii) *C_* is strictly increasing in ξ, and $C_* \to b$ as $\xi \to \infty$;*

(iii) *C_* is strictly decreasing in N, and $C_* \to 0$ as $N \to \infty$.* ■

Notice that C_* does not converge to 0 as $r \to 0$. Rather, it converges to the optimal cutoff for the undiscounted case. It does, of course, converge to the break-even point b as $r \to \infty$.

These comparative statics results for u_* and C_* can be derived directly from the explicit formulae given in Remark 4. A better approach, however, is to derive them from more general qualitative considerations. For example, the fact that C_* is strictly increasing in ξ — which is not completely obvious — can be derived as follows.

Note first that

$$u_*\left(C_*\left(\xi\right),\xi\right)-s=\frac{\xi\left(s-w\left(C_*\left(\xi\right)\right)\right)}{N}$$

by (5). Hence

$$\frac{dC_*}{d\xi}=\frac{\dfrac{s-w(C_*)}{N}-\dfrac{\partial u_*\left(C_*,\xi\right)}{\partial\xi}}{\dfrac{\partial u_*\left(C_*,\xi\right)}{\partial p}+\dfrac{\xi(\ h-\ell)}{N}}.$$

In other words, increasing ξ has a direct and an indirect effect. The direct effect is captured by the term $\frac{s-w(C_*)}{N}$ in the numerator. According to this effect, increasing ξ will tend to raise C_* and thereby lower experimentation. This is a free-rider effect. The indirect effect is captured by the term $-\frac{\partial u_*(C_*,\xi)}{\partial\xi}$ in the numerator. According to this effect, increasing ξ will tend to raise u_*. This in turn reduces C_*, and thereby raises experimentation. This is an encouragement effect.

Secondly, differentiating (1) with respect to ξ and using the envelope theorem, we obtain

$$\frac{\partial u_*}{\partial\xi}=\frac{1}{r}\Sigma\frac{u_*''}{2}+\frac{1}{r}(\xi+n)\frac{\Sigma}{2}\left(\frac{\partial u_*}{\partial\xi}\right)''.$$

In other words, $\frac{\partial u_*}{\partial\xi}$ is the expectation of the present discounted value of the shadow price of experimentation $\frac{1}{r}\Sigma\frac{u_*''}{2}$. Moreover (4) implies that

$$\frac{1}{r}\Sigma\frac{u_*''}{2}=\left\{\begin{array}{ll}\dfrac{u_*-s}{\xi} & \text{if}\quad p<C_*\\ \dfrac{u_*-w}{\xi+N} & \text{if}\quad p>C_*\end{array}\right\}.$$

Now, u_*-s is non-negative, convex, and $\left(u_*-s\right)(0)=0$. Similarly, u_*-w is non-negative, convex, and $\left(u_*-w\right)(1)=0$. Hence $\frac{1}{r}\Sigma\frac{u_*''}{2}$ is maximized at C_*, at which point it takes the value

$$\frac{u_*\left(C_*\left(\xi\right),\xi\right)-s}{\xi}=\frac{u_*\left(C_*\left(\xi\right),\xi\right)-w\left(C_*\left(\xi\right)\right)}{\xi+N}=\frac{s-w\left(C_*\left(\xi\right)\right)}{N}.$$

Hence $\frac{\partial u_*}{\partial\xi}<\frac{s-w(C_*(\xi))}{N}$ and $\frac{dC_*}{d\xi}>0$. In other words, the direct effect outweighs the indirect effect, and experimentation falls with the quality of background information.

7 The Free-Rider Effect and the Encouragement Effect

We begin our analysis of strategic experimentation by considering how a player's best response varies with variations in the strategies of the other players. For this purpose, it suffices to consider the case of a single player facing non-uniform background information. Indeed, from the point of view of any given individual, the background information and the experimentation of the other players can be regarded as non-uniform background information.

Suppose accordingly that the variation of background information with p is described by the function $\Xi : [0,1] \to (0,\infty)$, and suppose for concreteness that

$$\Xi(p) = (1-\beta)\,\underline{\Xi}(p) + \beta\,\overline{\Xi}(p),$$

where $\overline{\Xi} \geq \underline{\Xi}$. Let $u_*(p;\beta)$ denote the player's value function and let

$$\alpha_* = \frac{u_* - s}{s - w} - \Xi.$$

Then, applying the appropriate analogue of (5), we find that the player's best response is given by:

$$n_*(p;\beta) = \left\{ \begin{array}{ccc} = 0 & \text{if} & \alpha_* < 0 \\ \in \{0,1\} & \text{if} & \alpha_* = 0 \\ = 1 & \text{if} & \alpha_* > 0 \end{array} \right\}.$$

In other words, we may regard α_* as the player's incentive to experiment. Now

$$\frac{\partial \alpha_*}{\partial \beta} = -\left(\overline{\Xi} - \underline{\Xi}\right) + \frac{1}{s-w}\frac{\partial u_*}{\partial \beta}.$$

In other words, changes in β have two effects. The first is a free-rider effect: extra background information, which is supplied free, is used as a substitute for information which would otherwise have had to be supplied at an opportunity cost. The second is an encouragement effect: supplying extra information encourages the player to increase experimentation in order to bring forward the time at which the extra information becomes available.

An illustrative example may help to get a better grasp of the latter effect. Suppose that the player is initially indifferent between experimenting or not

given his prior belief p, and that there is no background information at all. Suppose now that he is told that another player will begin to experiment in the future whenever the prior moves above the break-even point. The fact that another player may contribute more information in the future gives a strict incentive to experiment to the initial player, since by experimenting the player now gets strictly more information than if the other player was not around. Indeed, if he did not experiment the prior would not move — since there is no background information — and he would not be able to benefit from the potential information provided by the future experimentation of the other player.

8 Strategic Experimentation: The Undiscounted Case

This section deals with the special case of our model in which there is no discounting. This case is easy to analyze. Indeed, it is possible to get an explicit characterization of both the team and the strategic solutions. In particular, we obtain a clear idea of the consequences of free riding for the strategic solution. This simplicity is, unfortunately, obtained at a cost: in the absence of discounting, there is no encouragement effect.

When there is no discounting, players evaluate outcomes according to the overtaking criterion. In order to understand what this involves, note first that, by exploiting the background information, player i can always ensure that her long-run average payoff

$$\lim_{T\to\infty} \mathrm{E}\left[\frac{1}{T}\int_0^T d\pi_i(t)\right]$$

is equal to the full-information payoff $u_0(p_0) = (1-p_0)s + p_0 h$. For example, she can choose the risky arm iff p exceeds the break-even probability b. Her experimentation strategy will therefore be chosen to maximize her transient payoff

$$\lim_{T\to\infty} \left(\mathrm{E}\left[\int_0^T d\pi_i(t)\right] - T\,u_0(p_0)\right).$$

Lemma 7 *Player i's transient payoff exists and is equal to*

$$\mathrm{E}\left[\int_0^\infty \left((1-x_i)\,s + x_i w(p(t)) - u_0(p(t))\right) dt\right]. \tag{6}$$

Note that player i's transient payoff is non-positive, and that it may be negatively infinite.

Proof. We have $\mathrm{E}\left[d\pi_i(t)\right] = (1-x_i)\,s + x_i w(p(t))$, where $x_i \in \{0,1\}$ is player i's action at time t, and $\mathrm{E}\left[u_0(p(t))\right] = u_0(p_0)$. Hence

$$\mathrm{E}\left[\int_0^T d\pi_i(t)\right] - T u_0(p_0) = \mathrm{E}\left[\int_0^T \left((1-x_i)\,s + x_i w(p(t)) - u_0(p(t))\right) dt\right].$$

But

$$(1-x_i)\,s + x_i w(p(t)) - u_0(p(t)) \leq 0.$$

Letting $T \to \infty$ and applying the monotone convergence theorem, we conclude that player i's transient payoff exists. (It may be negatively infinite.) ■

8.1 The Team Problem

Like the discounted team problem, the undiscounted team problem can be viewed as a controlled-variance problem. It can therefore be handled using dynamic programming.

Lemma 8 *The value function $v_* : [0,1] \to R$ for the undiscounted team problem is the unique bounded solution of the Bellman equation*

$$0 = \max_{n \in \{0,1,\ldots,N\}} \left(\left(\frac{N-n}{N}\, s + \frac{n}{N}\, w - u_0 \right) + (\xi + n)\, \Sigma \frac{v_*''}{2} \right). \tag{7}$$

Proof. Recall that player i's transient payoff takes the form

$$\mathrm{E}\left[\int_0^\infty \left((1-x_i)\,s + x_i w(p(t)) - u_0(p(t))\right) dt\right].$$

The objective of the team is therefore to maximize the undiscounted sum of the flow payoff

$$\frac{1}{N} \sum_{i=1}^{N} \left((1-x_i)\,s + x_i w(p) - u_0(p)\right)$$

$$= \frac{N-n}{N}\, s + \frac{n}{N}\, w\,(p) - u_0(p(t)),$$

where $n = \sum_{i=1}^{N} x_i$. Now, if the current belief is p, and if n players experiment, then the change in beliefs

$$dp \sim \mathrm{N}\left[0, (\xi + n)\,\Sigma\,(p)\,dt\right]$$

as before. Arguing as in the proof of Lemma 2, then, we find that the value function $v_* : [0,1] \to R$ for the undiscounted team problem is the unique bounded solution of (7). ■

Furthermore, arguing as in the proof of Lemma 3, we arrive at a randomized-stopping formulation for (7).

Lemma 9 *The value function $v_* : [0,1] \to R$ for the undiscounted team problem is the unique bounded solution of the Bellman equation*

$$0 = \max_{n \in \{0,1,\ldots,N\}} \left(-\frac{(s-w)}{N} + \frac{1}{\xi + n} \left(\frac{\xi\,(s-w)}{N} - (u_0 - s) \right) + \Sigma \frac{v_*''}{2} \right). \quad \blacksquare \tag{8}$$

Now (8) implies that the optimal team policy in the undiscounted case is to put

$$n_* = \left\{ \begin{array}{ll} 0 & \text{if} \quad u_0 - s < \dfrac{\xi}{N}(s-w) \\ \\ N & \text{if} \quad u_0 - s > \dfrac{\xi}{N}(s-w) \end{array} \right\}.$$

Given that $u_0 = (1-p)s + ph$ and $w = (1-p)\ell + ph$, we see at once that the team cutoff

$$C_* = \frac{\xi\,(s-\ell)}{N\,(h-s) + \xi(h-\ell)}. \tag{9}$$

It is notable just how much simpler the analysis has become in the absence of discounting: the optimal cutoff can be calculated without first solving for the value function v_*. The same short cut works in the strategic case, enabling us easily to characterize the set of Markov equilibria.

8.2 The Strategic Problem

We assume for simplicity that there are only two players.

Definition 1 *A* **mixed (Markov) strategy** *for player i is a mapping $X_i : [0,1] \to [0,1]$.*

A mixed strategy for player i specifies the probability with which he will choose to experiment in each state. (Pure strategies are the special case of mixed strategies in which players are required to experiment with probability 0 or 1 in all states.)

We discuss mixed strategies more extensively in the next section. Here we simply note that, when the player j employs the mixed strategy X_j, the Bellman equation for the value function $v_i(\cdot)$ of player i takes the form

$$0 = \max_{X_i \in [0,1]} \left(((1 - X_i)\ s + X_i w - u_0) + (\xi + X_j + X_i)\ \Sigma \frac{v_i''}{2} \right), \tag{10}$$

and that this equation can be reformulated as

$$0 = \max_{X_i \in [0,1]} \Bigg(- (s - w) + \frac{1}{\xi + X_j + X_i} \left((\xi + X_j)(s - w) - (u_0 - s) \right) + \Sigma \frac{v_i''}{2} \Bigg). \tag{11}$$

Equation (10) is identical to equation (7), except that N has been replaced by 1, ξ has been replaced by $\xi + X_j$ and x_i has been replaced by X_i. Similarly, equation (11) is identical to equation (8), except that N has been replaced by 1, ξ has been replaced by $\xi + X_j$ and x_i has been replaced by X_i.

The set of equilibria of our game can be characterized in terms of the incentive to experiment

$$\alpha_0 = \frac{u_0 - s}{s - w} - \xi.$$

Theorem 10 *The pair of mixed strategies (X_1, X_2) constitutes an equilibrium iff*

$$(X_1, X_2) \quad \left\{ \begin{array}{cll} = (0,0) & if & \alpha_0 < 0 \ and \ p < b \\ \in \{(0,1),(1,0),(\alpha_0,\alpha_0)\} & if & \alpha_0 \in [0,1] \ \ and \ p < b \\ = (1,1) & if & \alpha_0 > 1 \ or \ p \geq b \end{array} \right\}.$$

Notice that α_0 increases strictly from $-\xi$ to $+\infty$ as p increases from 0 to b. Hence there are unique cutoffs $0 < C_1 < C_2 < b$ such that $\alpha_0 = 0$ at C_1 and $\alpha_0 = 1$ at C_2. It is easy to check that

$$C_1 = \frac{\xi\,(s-\ell)}{(h-s)+\xi(h-\ell)} \tag{12}$$

and

$$C_2 = \frac{(\xi+1)\,(s-\ell)}{(h-s)+(\xi+1)\,(h-\ell)}. \tag{13}$$

In particular, comparing (12) with (9), we see that $C_* < C_1$.

In any pure-strategy equilibrium, neither player chooses the risky action when $p < C_1$, exactly one player chooses the risky action when $p \in (C_1, C_2)$, and both players choose the risky action when $p > C_2$. In particular, compared with the socially efficient outcome, two players too few experiment when $p \in (C_*, C_1)$, and one player too few experiments when $p \in (C_1, C_2)$. Because of free riding, equilibrium experimentation in any pure-strategy equilibrium is strictly less than would be socially efficient.

There are many pure-strategy equilibria. In one equilibrium, player 1 chooses the risky action for all $p \in (C_1, C_2)$ and player 2 free rides on player 1's experimentation in this range. In another equilibrium, the roles of the players are reversed. We refer to these two extreme asymmetric equilibria as pioneer-follower equilibria. More generally, from any partition of the interval (C_1, C_2) into two (measurable) subsets $S_1$and S_2, one can construct a pure-strategy equilibrium such that player 1 puts $x_1 = 1$ and player 2 puts $x_2 = 0$ when $p \in S_1$ and vice versa when $p \in S_2$. All these pure-strategy equilibria induce the same amount of total experimentation. They differ only in the allocation of the costs of experimentation between the two players.

There is, however, only one symmetric equilibrium. In this equilibrium, the players experiment with probability 0 for $p \leq C_1$, with probability rising continuously from 0 to 1 as p rises from C_1 to C_2, and with probability

1 for $p \geq C_2$. As in the pure-strategy equilibria, then, there is too little experimentation for $p < C_2$. Notice, however, that the symmetric equilibrium involves more experimentation than the pure-strategy equilibria when p is near C_2, and less experimentation than the pure-strategy equilibria when p is near C_1. Hence, in order to maximize experimentation, exactly one player should experiment for $\alpha_0 \in \left[0, \frac{1}{2}\right)$ and both players should mix for $\alpha_0 \in \left(\frac{1}{2}, 1\right]$.

Proof. Consider equation (11). When $p \geq b$, $(s - w) \leq 0$ and $(u_0 - s) \geq 0$ with at least one strict inequality, so that it is a dominant strategy for player i to choose $X_i = 1$. When $p < b$, equation (11) reduces to

$$0 = -(s - w) + (s - w) \max_{X_i} \left(\frac{X_j - \alpha_0}{\xi + X_j + X_i} \right) + \Sigma \frac{v_i''}{2}.$$

This tells us that player i should experiment if $X_j < \alpha_0$ and play safe if $X_j > \alpha_0$, and that she will be indifferent between the two actions if $X_j = \alpha_0$. More explicitly, for $p < b$, we have

$$(X_1, X_2) \quad \left\{ \begin{array}{lll} = (0,0) & \text{if} & \alpha_0 < 0 \\ \in \{(0,1),(1,0)\} & \text{if} & 0 \leq \alpha_0 \leq 1 \\ = (1,1) & \text{if} & 1 < \alpha_0 \end{array} \right\},$$

as required. ■

9 Mixed Strategies

Play in any given period of a discrete-time stochastic game involves a well established order of events. At the beginning of the period, players simultaneously and independently choose possibly random actions. They then obtain payoffs depending on the current state and the realized profile of actions. Finally, a new state is chosen according to a distribution that depends on the current state and the realized profile of actions. In continuous time, it is possible to establish an analogous order of events, and thereby to attach a meaning to randomization. This is the approach that we adopt in the present section.

Suppose accordingly that the current belief is $p \in [0, 1]$ and that player i uses the mixed action X_i. Suppose too that, in any given period:

(i) each player i chooses $x_i = 0$ with probability $1 - X_i$ and $x_i = 1$ with probability X_i, the choices being made simultaneously and independently;

(ii) each player i receives the payoff $r\left(\left(1 - x_i\right) s + x_i \mu\right) dt$;

(iii) the background signal $d\pi_0(t) = \sqrt{\xi}\mu\, dt + \sigma\, dZ_0(t)$ and the payoffs $d\pi_i(t) = \left(\left(1 - x_i\right) s + x_i\mu\right) dt + \sigma\, dZ_i(t)$ are realized;

(iv) the players all observe the actions x_i for $1 \leq i \leq N$ and the payoffs $d\pi_i(t)$ for $0 \leq i \leq N$;

(v) a new belief $p + dp$ is generated.

Suppose finally that player i's continuation payoffs are given by $c_i : [0,1] \to R$.

Then, arguing as in Sections 3 and 4, one would expect that the expectation of player i's payoff conditional on the profile of actions $x = \times_{i=1}^{N} x_i$ would be

$$r\, dt \left(\left(\left(1 - x_i\right) s + x_i\, w(p)\right) + \frac{1}{r}\left(\xi + \sum_{j=1}^{N} x_j\right)\Sigma(p)\,\frac{c_i''(p)}{2} - c_i(p)\right) + c_i(p).$$

The unconditional expectation of her payoff would therefore be

$$H_i\left(X_i, X_{-i}, p, c_i(\cdot)\right) =$$

$$r\, dt \left(\left(\left(1 - X_i\right) s + X_i\, w(p)\right) + \frac{1}{r}\left(\xi + \sum_{j \neq i} X_j + X_i\right)\Sigma(p)\,\frac{c_i''(p)}{2} - c_i(p)\right) + c_i(p),$$

where $X_{-i} = \times_{j \neq i} X_j$. Finally, her value function $u_i(\cdot)$ when the other players use the mixed-strategy profile X_{-i} would be the unique bounded solution of the Bellman equation

$$u_i(p) = \max_{X_i \in [0,1]} H_i\left(X_i, X_{-i}(p), p, c_i(\cdot)\right) \text{ for all } p \in [0,1].$$

We are therefore led to the following definition.

Definition 2 *The value function $u_i : [0,1] \to R$ for player i when the other players use the mixed-strategy profile X_{-i} is the unique bounded solution of the Bellman equation*

$$u_i = \max_{X_i \in [0,1]} \left(\left((1 - X_i)\, s + X_i\, w\left(p\right) \right) + \frac{1}{r} \left(\xi + \sum_{j \neq i} X_j + X_i \right) \Sigma\left(p\right) \frac{u_i''(p)}{2} \right). \tag{14}$$

Several aspects of this definition are worthy of comment. First, in the case where the players other than i use pure strategies and player i confines herself to best responses that are pure strategies, equation (14) can be derived instead of assumed. More explicitly, the value function $u_i\left(\cdot\right)$ of player i when the other players use the pure-strategy profile $x_{-i} = \times_{j \neq i} x_j$ can be shown to be the unique bounded solution of the Bellman equation

$$u_i = \max_{x_i \in \{0,1\}} \left(\left((1 - x_i)\, s + x_i\, w\left(p\right) \right) + \frac{1}{r} \left(\xi + \sum_{j \neq i} x_j + x_i \right) \Sigma\left(p\right) \frac{u_i''(p)}{2} \right). \tag{15}$$

This equation should be compared with equation (1). The only difference between the two equations is that, in (15), N has been replaced by 1 and ξ has been replaced by $\xi + \sum_{j \neq i} x_j$.

Secondly, although we have chosen to assume rather than derive equation (14), we believe that it is in principle possible to derive it. To do so, it would be necessary first to build a mathematical framework within which the mixed extension of our game can be formulated, and then to derive (14) from this formulation. This is, in effect, a more elaborate version of the old problem of finding a mathematical framework within which it is possible to derive the result that the empirical distribution of a continuum of i.i.d. random variables is equal to the population distribution with probability one. Building such a mathematical framework is beyond the scope of the present paper.

Thirdly, although we ultimately assume equation (14), we have offered a heuristic derivation. We hope that this derivation will help to motivate our definition.

Fourthly, note that equation (14) differs from equation (15) only in that x_j has been replaced with X_j for all j. This is a result of the additive separability of the payoff of player i in the x_j. This additive separability

turns out to have a second consequence: the mixed extension of our game is strategically equivalent to the public-randomization extension of our game. Hence the symmetric mixed-strategy equilibria that we analyze below can also be interpreted as symmetric public-randomization equilibria in which players either all experiment or all play safe. See Section 12 below.

Finally, just as equation (1) has the alternative randomized-stopping formulation (4), so (14) has the randomized-stopping formulation given in the following lemma.

Lemma 11 *The value function $u_i : [0,1] \to R$ for player i when the other players use the mixed-strategy profile X_{-i} is the unique bounded solution of the Bellman equation*

$$0 = \max_{X_i \in [0,1]} \left(-r(s-w) + \frac{r}{\xi + \sum_{j \neq i} X_j + X_i} \left(\left(\xi + \sum_{j \neq i} X_j \right)(s-w) - (u_i - s) \right) + \Sigma \frac{(u_i - s)''}{2} \right). \quad \blacksquare \tag{16}$$

Note that equation (16) is identical to equation (4) except in that N has been replaced by 1, ξ has been replaced by $\xi + \sum_{j \neq i} X_j$ and x_i has been replaced by X_i.

Remark 5 *According to the perspective on randomization in continuous-time developed in this section, play of a period of a continuous-time stochastic game involves exactly the same sequence of events as play of a period of a discrete-time stochastic game. It is also true that, in both cases, there is a well defined first period, and that the set of periods is totally ordered. The difference between the two cases consists in the fact that, in continuous time, the set of periods is not well ordered. In particular, in continuous time, it is not the case that there is a well defined period immediately following any other period. It is not therefore possible to build up the path that results from a profile of strategies inductively.*

Remark 6 *It is possible to give an alternative interpretation to the strategies X_i. According to this interpretation, the players divide their time between*

experimenting and not experimenting. Thus X_i is the proportion of time that player i devotes to experimenting, and $1 - X_i$ is the proportion of time she devotes to not experimenting. See Harris [12].

10 Strategic Experimentation: the Discounted Case

In this section we tackle the discounted case. This case is harder to analyze than the undiscounted case, but it exhibits the encouragement effect. We focus throughout on symmetric equilibria.

Our first objective is to obtain a partial characterization of equilibrium experimentation. To this end, we prove the following lemma, which is a simple generalization of Lemma 4.

Lemma 12 *Suppose that player j chooses the mixed strategy X_j for all $j \neq i$, and that player i plays a best response to $\times_{j \neq i} X_j$. Let u_i denote player i's value function. Then, for all $p \in (0,1)$:*

(i) $u_\infty(p) < u_i(p) < u_0(p)$;

(ii) $u_i''(p) > 0$.

Proof. Let $u_*(\cdot; r, N, \xi)$ denote the value function for the team problem with discount rate r, N players and background information ξ. Then $u_i \geq u_*(\cdot; r, 1, \xi)$ since $\sum_{j \neq i} X_j \geq 0$, and $u_i \leq u_*(\cdot; r, 1, \xi + N - 1)$ since $\sum_{j \neq i} X_j \leq N - 1$. Part (i) therefore follows from part (i) of Lemma 4. Next, (16) can be written in the form

$$\frac{1}{r} \Sigma \frac{u_i''}{2} = \min \left\{ \frac{u_i - s}{\xi + \sum_{j \neq i} X_j}, \frac{u_i - w}{\xi + \sum_{j \neq i} X_j + 1} \right\}.$$

Part (ii) therefore follows from part (i). ■

The partial characterization of equilibrium experimentation is then contained in the following theorem, which should be compared with Theorem 10.

Theorem 13 *Suppose that $X_\dagger$ is a symmetric equilibrium, and let $u_\dagger$ be the associated value function. Let*

$$\alpha_\dagger = \frac{u_\dagger - s}{s - w} - \xi.$$

Then

$$X_\dagger = \left\{ \begin{array}{ccc} 0 & \text{if} & \alpha_\dagger < 0 \text{ and } p < b \\ \dfrac{\alpha_\dagger}{N-1} & \text{if} & \alpha_\dagger \in [0, N-1] \text{ and } p < b \\ 1 & \text{if} & \alpha_\dagger > N-1 \text{ or } p \geq b \end{array} \right\}.$$

Note that Lemma 12 shows that $\alpha_\dagger$ increases strictly from $-\xi$ to $+\infty$ as p increases from 0 to b. Hence there are unique cutoffs $0 < C_1 < C_N < b$ such that $\alpha_\dagger = 0$ at C_1 and $\alpha_\dagger = N - 1$ at C_N. In particular, equilibrium experimentation $X_\dagger$ is 0 for $p \in [0, C_1]$, increases strictly from 0 to 1 as p increases from C_1 to C_N, and is 1 for $p \geq C_N$. Lemma 12 even shows that $\alpha_\dagger$ is strictly convex on $[0, b]$. Hence $X_\dagger$ is convex on $[0, C_N]$, and strictly convex on $[C_1, C_N]$.

Note too that it follows from (5) that the team cutoff C_* is the unique solution of the equation

$$\frac{u_* - s}{s - w} - \frac{\xi}{N} = 0.$$

Moreover $u_* \geq u_\dagger$ (because a team can always undertake the same amount of experimentation as is undertaken in equilibrium), and $\frac{\xi}{N} < \xi$. Hence $C_* < C_1$. In particular, because of free riding, equilibrium experimentation is strictly less than would be socially efficient.

Note finally that the characterization of equilibrium experimentation $X_\dagger$ given in Theorem 13 is partial because it depends on $u_\dagger$, which is itself endogenous. (We do, of course, know several properties of $u_\dagger$.) This contrasts with the characterization of equilibrium experimentation given in Theorem 10, which was completely explicit since it depended only on u_0.

Proof. $X_\dagger$ is a symmetric equilibrium iff

$$X_\dagger \in \underset{X_i \in [0,1]}{\operatorname{argmax}} \left(\frac{r}{\xi_\dagger + X_i} \left(\frac{\xi_\dagger (s - w)}{N} - (u_\dagger - s) \right) \right), \tag{17}$$

where $u_\dagger$ is the unique bounded solution of the equation

$$0 = \max_{X_i \in [0,1]} \left(-r\left(s-w\right) + \frac{r}{\xi_\dagger + X_i} \left(\frac{\xi_\dagger \left(s-w\right)}{N} - \left(u_\dagger - s\right) \right) + \Sigma \frac{\left(u_\dagger - s\right)''}{2} \right) \tag{18}$$

and

$$\xi_\dagger = \xi + \left(N-1\right) X_\dagger. \tag{19}$$

Now suppose that $p < b$. Then there are three possibilities.

(i) If $u_\dagger - s < \xi_\dagger \left(s-w\right)$ then $X_\dagger = 0$ from (17). Hence $\xi_\dagger = \xi$ from (19), and $\alpha_\dagger = \frac{u_\dagger - s}{s-w} - \xi = \frac{u_\dagger - s}{s-w} - \xi_\dagger = \frac{u_\dagger - s - \xi_\dagger (s-w)}{s-w} < 0$.

(ii) If $u_\dagger - s = \xi_\dagger \left(s-w\right)$ then $\xi_\dagger = \frac{u_\dagger - s}{s-w} = \alpha_\dagger + \xi$, and $X_\dagger = \frac{\xi_\dagger - \xi}{N-1}$ from (19). Hence $X_\dagger = \frac{\alpha_\dagger}{N-1}$ and $\alpha_\dagger = \left(N-1\right) X_\dagger \in [0, N-1]$.

(iii) If $u_\dagger - s > \xi_\dagger \left(s-w\right)$ then $X_\dagger = 1$ from (17). Hence $\xi_\dagger = \xi + N - 1$ from (19), and $\alpha_\dagger = \frac{u_\dagger - s}{s-w} - \xi = \frac{u_\dagger - s}{s-w} - \xi_\dagger + N - 1 = \frac{u_\dagger - s - \xi_\dagger (s-w)}{s-w} + N - 1 > N - 1$.

If, on the other hand, $p \geq b$, then Lemma 12 implies that $s - w \leq 0$ and $u_\dagger - s \geq 0$ with at least one strict inequality. Hence $X_\dagger = 1$ from (17). ■

By developing Theorem 13 a little further, we obtain a characterization of the set of value functions of symmetric equilibria which does not involve any reference to the associated strategies. This characterization is of intrinsic interest. It could also form the basis for an approach to proving the uniqueness of equilibrium.

For all p, u such that $p \in [0,1]$ and $u \in [u_\infty (p), u_0 (p)]$, let

$$A\left(p,u\right) = \frac{u-s}{s - w\left(p\right)} - \xi,$$

and let

$$T\left(p,u\right) = \left\{ \begin{array}{ccc} 0 & \text{if} & A\left(p,u\right) < 0 \text{ and } p < b \\ \dfrac{N A\left(p,u\right)}{N-1} & \text{if} & A\left(p,u\right) \in [0, N-1] \text{ and } p < b \\ N & \text{if} & A\left(p,u\right) > N-1 \text{ or } p \geq b \end{array} \right\}.$$

Here $A\left(p,u\right)$ should be thought of as the incentive to experiment when the belief is p and the continuation payoff is u, and $T\left(p,u\right)$ should be thought

of as total experimentation in the same case. Then we have the following corollary of Theorem 13.

Corollary 14 *$u_\dagger$ is the value function of a symmetric equilibrium iff $u_\dagger$ is a bounded solution of*

$$0 = -\frac{r(s-w)}{N} + \frac{r}{\xi + T(u_\dagger)}\left(\frac{\xi(s-w)}{N} - (u_\dagger - s)\right) + \Sigma\frac{(u_\dagger - s)''}{2}, \quad (20)$$

where we have suppressed the dependence of T on p.

The characterization the set of value functions of symmetric equilibria contained in Corollary 14 allows us to find the set of all symmetric equilibria by means of a two-step procedure: first find all solutions $u_\dagger$ of (20), then calculate the associated strategies $X_\dagger$. More explicitly, we have the following second corollary to Theorem 13.

Corollary 15 *$X_\dagger$ is a symmetric equilibrium iff there exists a solution $u_\dagger$ of (20) such that $X_\dagger(p) = T(p, u_\dagger(p))$ for all $p \in [0,1]$.*

In particular, a good approach to the uniqueness of equilibrium would be to show that (20) has a unique solution.

Proof of Corollary 14. Suppose that $u_\dagger$ is the value function of the symmetric equilibrium $X_\dagger$. Combining (17), (18) and (19), we have

$$0 = -r(s-w) + \frac{r}{\xi + NX_\dagger}\left(\left(\xi + (N-1)X_\dagger\right)(s-w) - (u_\dagger - s)\right) + \Sigma\frac{(u_\dagger - s)''}{2}. \quad (21)$$

Writing $\xi + (N-1)X_\dagger$ in the form $\frac{\xi}{N} + \frac{N-1}{N}(\xi + NX_\dagger)$, we get

$$0 = -\frac{r(s-w)}{N} + \frac{r}{\xi + NX_\dagger}\left(\frac{\xi(s-w)}{N} - (u_\dagger - s)\right) + \Sigma\frac{(u_\dagger - s)''}{2}.$$

Moreover $NX_\dagger = T(u_\dagger)$ by Theorem 13. Hence (20) is satisfied.

Suppose now that $u_\dagger$ is a bounded solution of (20), and put $X_\dagger = \frac{T(u_\dagger)}{N}$. Then, reversing the argument of the previous paragraph, we see that (21) holds. Moreover $X_\dagger = 0$ if $u_\dagger - s < \xi(s-w)$ and $p < b$, and $X_\dagger = 1$ if $u_\dagger - s > \xi(s-w)$ and $p < b$. That is, $X_\dagger$ is a best response for player i when

the other players choose $X_\dagger$, and when player i is constrained to experiment whenever $p \geq b$. To complete the proof, then, we need only show that the constraint on experimentation for $p \geq b$ is not binding. This follows at once from Lemma 12, which implies that , for all optimal strategies X_* in the unconstrained problem, $X_* = 1$ for $p \geq b$. ■

Proof of Corollary 15. The corollary follows immediately from Theorem 13 and Corollary 14. ■

We have already seen that our experimentation game is not a supermodular game: even in a two-player game, increasing the level of experimentation of one player in all states may lead the other player to increase experimentation in some states and to decrease experimentation in other states. This is because, roughly speaking, current experimentation by one player is a strategic substitute for current experimentation by another, but future experimentation by one player is a strategic complement for current experimentation by another.

As a result, we cannot establish existence by applying Tarski's fixed-point theorem to the best-response mapping. We can, however, exploit the encouragement effect to arrive at an increasing mapping to which we can apply Tarski's theorem. More precisely, we know that

- a player's payoff from her best response is increasing in the total experimentation of the other players.

Moreover Theorem 13 tells us that

- a player's level of experimentation is increasing in her payoff.

We may therefore construct an increasing map in the space of value functions. Applying Tarski's fixed-point theorem to this mapping, we can establish the existence of a maximum symmetric equilibrium.

Note that there are some grounds for believing that our model possesses a *unique* symmetric equilibrium. We do not, however, have a complete proof of this, and we must therefore admit the possibility of multiple equilibria.

Definition 3 *Suppose that u and $\widehat{u}$ are the value functions of two symmetric equilibria. Then the first equilibrium* **dominates** *the second if $u(p) \geq \widehat{u}(p)$ for all $p \in [0,1]$.*

Definition 4 *An equilibrium is* **maximum symmetric** *if it is symmetric and it dominates all other symmetric equilibria.*

We use the terminology 'maximum' rather than 'maximal' in order to stress that the equilibrium that we find dominates all other equilibria rather than simply being undominated.

Theorem 16 *There exists a maximum symmetric equilibrium.*

We denote the value function of a representative player in the maximum symmetric equilibrium by $u_{\#}$.

Proof. For all $p \in [0,1]$ and all $u \in [u_{\infty}(p), u_0(p)]$, let

$$\tau(p,u) = \left\{ \begin{array}{ccc} \xi & \text{if} & A(p,u) < 0 \text{ and } p < b \\ \xi + A(p,u) & \text{if} & A(p,u) \in [0, N-1] \text{ and } p < b \\ \xi + N - 1 & \text{if} & A(p,u) > N-1 \text{ or } p \geq b \end{array} \right\}.$$

Here $\tau(p,u)$ should be thought of as the total information available to a player — i.e. as background information plus total experimentation by the other players — when the belief is p and the continuation payoff is u.

Let $\mathcal{U}$ denote the set of functions $u : [0,1] \to [s,h]$ such that: (i) $u_{\infty} \leq u \leq u_0$; and (ii) u is convex. For all $u \in \mathcal{U}$, let $\phi_1(u) : [0,1] \to [0, \xi + N - 1]$ be defined by $(\phi_1(u))(p) = \tau(p, u(p))$ for all $p \in [0,1]$, and for all $\Xi : [0,1] \to [0, \xi + N - 1]$, let $\phi_2(\Xi) \in \mathcal{U}$ be the value function of the one-player when background information is Ξ. Finally, let $\phi = \phi_2 \circ \phi_1$. Then it is easy to see that: (i) $\mathcal{U}$ is closed under the operation of taking pointwise maxima; (ii) ϕ is a self-map of $\mathcal{U}$; and (iii) ϕ is non-decreasing in the sense that $u \geq \widehat{u}$ pointwise implies that $\phi(u) \geq \phi(\widehat{u})$ pointwise. (Notice that $\mathcal{U}$ is not a lattice, since it is not closed under the operation of taking pointwise minima.) The proof of Tarski's fixed-point theorem therefore shows that ϕ possesses a maximum fixed point. We denote this fixed point by $u_{\#}$.

In order to show that there exists a maximum symmetric equilibrium, then, it suffices to show that $u_{\dagger}$ is the value function of a symmetric equilibrium iff $u_{\dagger}$ is a fixed point of ϕ.

Suppose accordingly that $u_{\dagger}$ is a fixed point of ϕ. Put $\Xi_{\dagger} = \phi_1(u_{\dagger})$ and $X_{\dagger} = \frac{\Xi_{\dagger} - \xi}{N-1}$. By construction, $u_{\dagger}$ is the value function of player i when all players $j \neq i$ employ $X_{\dagger}$, and player i chooses a best response. Hence, in order to show that $u_{\dagger}$ is the value function of a symmetric equilibrium, we

need only show that $X_\dagger$ is a best response for player i when all players $j \neq i$ employ $X_\dagger$. Put $\alpha_\dagger = \frac{u_\dagger - s}{s-w} - \xi$. If $p < b$ then:

(i) If $\alpha_\dagger < 0$ then $\Xi_\dagger = \xi$ by definition of ϕ_1. Hence $X_\dagger = 0$ and $u_\dagger - s = (\alpha_\dagger + \xi)(s-w) = (\alpha_\dagger + \Xi_\dagger)(s-w) < \Xi_\dagger (s-w)$.

(ii) If $\alpha_\dagger \in [0, N-1]$ then $\Xi_\dagger = \xi + \alpha_\dagger$ by definition of ϕ_1. Hence $X_\dagger = \frac{\alpha_\dagger}{N-1}$ and $u_\dagger - s = (\alpha_\dagger + \xi)(s-w) = \Xi_\dagger (s-w)$.

(iii) If $\alpha_\dagger > N-1$ then $\Xi_\dagger = \xi + N - 1$ by definition of ϕ_1. Hence $X_\dagger = 1$ and $u_\dagger - s = (\alpha_\dagger + \xi)(s-w) > (N-1+\xi)(s-w) > \Xi_\dagger (s-w)$.

If, on the other hand, $p \geq b$, then $u_\dagger - s \geq 0$ and $s - w \geq 0$ with at least one strict inequality, and therefore $u_\dagger - s > \Xi_\dagger (s-w)$. Overall, then, $X_\dagger = 0$ whenever $u_\dagger - s < \Xi_\dagger (s-w)$ and $X_\dagger = 1$ whenever $u_\dagger - s > \Xi_\dagger (s-w)$. That is, $X_\dagger$ is a best response to $X_\dagger$.

Suppose finally that $X_\dagger$ is a symmetric equilibrium, and let $u_\dagger$ be the associated value function. Put $\Xi_\dagger = \xi + (N-1) X_\dagger$. Then $u_\dagger = \phi_2 (\Xi_\dagger)$ by definition of ϕ_2. Moreover $\Xi_\dagger = \phi_1 (u_\dagger)$ by Theorem 13. ■

11 Comparative Statics of the Strategic Problem

In this section we examine the comparative statics of the maximum symmetric equilibrium with respect to the discount rate r, the number of players N and the level of background information ξ.

Theorem 17 *For all $p \in (0,1)$:*

(i) $u_\# (p)$ is strictly decreasing in r, $u_\# (p) \to u_0 (p)$ as $r \to 0$, and $u_\# (p) \to u_\infty (p)$ as $r \to \infty$;

(ii) $u_\# (p)$ is strictly increasing in N, and $u_\# (p)$ is bounded away from $u_0 (p)$ as $N \to \infty$;

(iii) $u_\# (p)$ is strictly increasing in ξ, and $u_\# (p) \to u_0 (p)$ as $\xi \to \infty$.

Notice that what is remarkable about the comparative-statics results of Theorem 17 is not their sign, but rather the fact that they hold at all in a strategic problem. Notice too that the full-information payoff is not obtained even in the limit as $N \to \infty$. This underlines the strength of the free-rider effect in our model.

Proof. The monotonicity results follow as corollaries of the existence construction. In the case of r: ϕ_1 is independent of r; and $(\phi_2(\Xi))(p)$ is strictly decreasing in r for $p \in (0,1)$. In the case of N: $(\phi_1(\Xi))(p)$ is non-decreasing in N, and strictly increasing for $p \in [b,1]$; and ϕ_2 is independent of N. In the case of ξ: $(\phi_1(\Xi))(p)$ is non-decreasing in ξ, and strictly increasing for $p \in [b,1]$; and ϕ_2 is independent of ξ.

The limit results in parts (i) and (iii) can be deduced from the corresponding results for the team problem. Let $u_*(\cdot; r, N, \xi)$ denote the value function for the team problem with discount rate r, N players and background information ξ. Since $u_*(p; r, 1, \xi) \leq u_\#(p) \leq u_0(p)$ and $u_*(\cdot; r, 1, \xi)$ converges uniformly to u_0 as $r \to 0$ and as $\xi \to \infty$, $u_\#$ too converges uniformly to u_0 as $r \to 0$ and as $\xi \to \infty$. Similarly, since $u_\infty(p) \leq u_\#(p) \leq u_*(p; r, 1, \xi + N - 1)$ and $u_*(\cdot; r, 1, \xi + N - 1)$ converges uniformly to u_∞ as $r \to \infty$, $u_\#$ too converges uniformly to u_∞ as $r \to \infty$.

As for the limit result in part (ii), we let C_1 be the unique solution of the equation $u_0(p) - s = \xi(s - w(p))$, as in (12) above. Since $u_\# \leq u_0$, we must have $X_\# = 0$ for $p < C_1$. Hence $u_\#(p) \leq u_*(p; r, 1, \Xi)$, where

$$\Xi = \left\{ \begin{array}{ccc} \xi & \text{if} & p < C_1 \\ \xi + N - 1 & \text{if} & p \geq C_1 \end{array} \right\},$$

and $u_*(\cdot; r, 1, \Xi)$ is the value function of the one-player problem with non-uniform background information Ξ. Now it is easy to check that $u_*(p; r, 1, \Xi)$ is non-decreasing in N, and that it converges to the unique bounded solution $\overline{u}$ of the equation

$$\left\{ \begin{array}{ccc} \frac{1}{r}\xi\Sigma\frac{\overline{u}''}{2} + s - \overline{u} = 0 & \text{if} & p < C_1 \\ \overline{u}'' = 0 & \text{if} & p \geq C_1 \\ \overline{u} = h & \text{if} & p = 1 \end{array} \right\}$$

uniformly as $N \to \infty$. It is also easy to check that $\overline{u} < u_0$ on $(0,1)$. ■

Further insight into the source of the monotonicity results can be obtained by considering the equation of variations for the value function of the maximum symmetric equilibrium. Indeed, the total information available in the maximum symmetric equilibrium to any given player due to background information and to experimentation by the other players is

$$\Xi = \max\left\{\xi, \min\left\{\xi + N - 1, \frac{u_{\#} - s}{s - w}\right\}\right\}$$

by Theorem 13. Hence the value function of this equilibrium must satisfy the Bellman equation

$$u_{\#} = \max_{X \in [0,1]} \left((1 - X)\, s + Xw + \frac{1}{r}\,(\Xi + X)\,\Sigma \frac{u''_{\#}}{2} \right).$$

Hence, applying the envelope theorem, we obtain

$$\frac{\partial u_{\#}}{\partial r} = f + \frac{1}{r}\,(\Xi + X)\,\frac{\Sigma}{2}\left(\frac{\partial u_{\#}}{\partial r}\right)'',$$

where

$$f = -\frac{1}{r^2}\,(\Xi + X)\,\Sigma \frac{u''_{\#}}{2} + \frac{1}{r}\frac{\partial \Xi}{\partial r}\frac{\Sigma}{2}\left(\frac{\partial u_{\#}}{\partial r}\right)''$$

and

$$\frac{\partial \Xi}{\partial r} = \left\{ \begin{array}{ccc} 0 & \text{if} & \alpha_{\#} < 0 \text{ and } p < b \\ \dfrac{1}{s - w}\dfrac{\partial u_{\#}}{\partial r} & \text{if} & \alpha_{\#} \in [0, N-1] \text{ and } p < b \\ 0 & \text{if} & \alpha_{\#} > N - 1 \text{ or } p \geq b \end{array} \right\}.$$

In other words, when r increases, there are two effects on a player's payoff:

(i) The direct effect on her objective of the increase in r.

(ii) The indirect effect on her objective of the change induced in the behavior of the other players by the change in r.

The first effect is negative: she now values the future less. The second effect is a strategic effect: the rise in r discourages experimentation by the other players when $\alpha_{\#} \in [0, N-1]$ and $p < b$, and this further lowers her payoff.

The effect on her objective of the change induced in her own behavior by the change in r is of course zero by the envelope theorem. Notice the positive feedback implicit in the second effect. This is yet another manifestation of the encouragement effect.

A similar analysis of the cases of N and ξ can be undertaken. Increasing N has no direct effect on a player's objective. However, it does raise experimentation by the other players through an encouragement effect when $\alpha_{\#} \in [0, N-1]$ and $p < b$, and through an increase in the total number of players experimenting when $\alpha_{\#} > N-1$ or $p \geq b$. It also lowers per capita experimentation by the other players through a free-rider effect when $\alpha_{\#} \in [0, N-1]$ and $p < b$, but this decrease is exactly offset by the increase in the number of players. Increasing ξ has a direct effect on the objective: more information is provided in every state. It also has both a free-rider and an encouragement effect when $\alpha_{\#} \in [0, N-1]$ and $p < b$. Fortunately, the direct effect exactly offsets the free-rider effect, so that the net effect when $\alpha_{\#} \in [0, N-1]$ and $p < b$ is precisely the encouragement effect.

Finally, note that there is a unique symmetric equilibrium $X_\dagger$ when $r = 0$. Total information in this equilibrium is

$$\xi + NX_\dagger = \left\{ \begin{array}{lll} \xi & \text{if} & \alpha_0 < 0 \text{ and } p < b \\ \xi + \dfrac{N\alpha_0}{N-1} & \text{if} & \alpha_0 \in [0, N-1] \text{ and } p < b \\ \xi + N - 1 & \text{if} & \alpha_0 > N-1 \text{ or } p \geq b \end{array} \right\},$$

where

$$\alpha_0 = \frac{u_0 - s}{s - w} - \xi.$$

From this it follows that total information is independent of N for $p < C_1$, strictly decreasing in N for $p \in [C_1, C_N]$ and strictly increasing in N for $p > C_N$, where

$$C_1 = \frac{\xi\,(s - \underline{w})}{(h - s) + \xi(h - \underline{w})}$$

and

$$C_N = \frac{(\xi + N - 1)\,(s - \underline{w})}{(h - s) + (\xi + N - 1)\,(h - \underline{w})}.$$

Similarly, total information is strictly increasing in ξ for $p < C_1$, strictly decreasing in ξ for $p \in [C_1, C_N]$ and strictly increasing in ξ for $p > C_N$.

In other words, total information is not in general monotonic either in the number of players or the quantity of background information. This underlines the subtlety of the comparative statics results given in Theorem 17.

12 Public Randomization

In this section we analyze the public-randomization extension of our game. This extension is motivated by the idea that one way of capturing the intuitively natural social-experimentation pattern of taking turns is to introduce a public-randomization device, and to have players coordinate their experimentation on the basis of the realization of this device in such a way that exactly one player experiments at any given time and each player experiments with equal probability. For example, two players might toss a coin, with player 1 experimenting if the coin comes up heads and player 2 experimenting if the coin comes up tails. It emerges that literally taking turns is not an equilibrium, but that there is a wide variety of equilibria in which players achieve equilibrium by coordinating on the basis of a public-randomization device instead of acting independently on the basis of a private-randomization device.

The most convenient way of characterizing the public-randomization equilibria of our model is to show that the mixed extension of our game is strategically equivalent to its public-randomization extension. All the symmetric mixed-strategy equilibria that we have analyzed in this paper are therefore also symmetric public-randomization equilibria, and vice versa.

It will be helpful to begin by reformulating the notion of a mixed strategy. For the purposes of this section, we shall take it that a mixed strategy for player i is a mapping $X_i : [0,1]^2 \to \{0,1\}$. Such a strategy specifies the action $X_i(p, \eta_i)$ that player i will take when the state is p and her private signal is η_i. We shall also take it that, in any given period of the mixed extension of our game, if the current state is p then:

(i) Nature chooses a profile $\eta = \times_{i=1}^{N} \eta_i$ of private signals, the signal η_i being distributed uniformly on $[0,1]$ and independently of all other past, current and future private signals;

(ii) each player i observes her private signal η_i;

(iii) each player i takes the action $x_i = X_i(p, \eta_i)$, the actions being chosen simultaneously;

(iv) each player i receives payoff $r((1 - x_i)\, s + x_i \mu)\, dt$;

and so on.

Similarly, a public-randomization strategy for player i is a mapping $X_i : [0,1]^2 \to \{0,1\}$. Such a strategy specifies the action $X_i(p, \zeta)$ that player i will take when the state is p and the public signal is ζ. We shall also take it that, in any given period of the public-randomization extension of our game, if the current state is p then:

(i) Nature chooses a public signal ζ, the signal ζ being distributed uniformly on $[0,1]$ and independently of all other past, current and future public signals;

(ii) each player i observes the public signal ζ;

(iii) each player i takes the action $x_i = X_i(p, \zeta)$, the actions being chosen simultaneously;

(iv) each player i receives payoff $r((1 - x_i)\, s + x_i \mu)\, dt$;

and so on.

It is immediate from these formulations of the mixed and public-randomization extensions of our game that a strategy $X_i : [0,1]^2 \to \{0,1\}$ can serve either as a mixed strategy or as a public-randomization strategy. Moreover it follows easily from the additive separability of payoffs in the stage game that a strategy profile $X = \times_{i=1}^{N} X_i$ yields the same profile of payoffs whether it is used as a mixed-strategy profile or as a public-randomization-strategy profile. In other words, the mixed extension of our game and its public-randomization extension are strategically equivalent.[1] In particular, our analysis of symmetric mixed-strategy equilibria applies equally well to symmetric public-randomization-strategy equilibria.

Translated into distributional terms, strategic equivalence implies that, to each symmetric mixed-strategy equilibrium in which each player independently chooses to play safe with probability $1 - X_\dagger$ and to experiment with

[1] Strategic equivalence applies to payoffs, but not to equilibrium paths: the equilibrium paths of the mixed extension and the public-randomization extension may differ.

probability $X_\dagger$, there corresponds a symmetric public-randomization-strategy equilibrium in which players collectively choose to play safe with probability $1-X_\dagger$ and to experiment with probability $X_\dagger$. The difference between the two equilibria is that in the former case the number of experiments is binomially distributed, taking the value n with probability $\binom{N}{n}(1-X_\dagger)^{N-n}X_\dagger^n$, and in the latter case the number of experiments is 0 with probability $1-X_\dagger$ and N with probability $X_\dagger$. Similarly, to each symmetric public-randomization-strategy equilibrium in which players collectively choose to play safe with probability $1-X_\dagger$ and to experiment with probability $X_\dagger$, there corresponds a symmetric mixed-strategy equilibrium in which each player independently chooses to play safe with probability $1-X_\dagger$ and to experiment with probability $X_\dagger$.

More generally, if $X_\dagger$ is a symmetric mixed-strategy equilibrium, then any joint distribution over action profiles whose marginal over the action set of player i assigns probability $1-X_\dagger$ to the safe action and probability $X_\dagger$ to the risky action is a public-randomization equilibrium, and vice versa. The point is that, because payoffs in the stage game are additively separable in actions, it is the total probability with which each player experiments that matters, not the correlation between the actions of the different players. C.f. Harris [12].

13 Generalization of the Theory

The theory that we have developed in this paper can be generalized in three basic directions:

(i) the risky arm can take on an arbitrary (finite) number of values;

(ii) choice of the risky arm can generate a whole vector of signals (which may or may not include the payoff);

(iii) players can be allowed to choose between two risky arms.

The purpose of the present section is to sketch a single model that incorporates all three of these possibilities, and to outline the corresponding generalization of the theory.

13.1 The General Model

As before, there are N risk-neutral players. At time t, each player i chooses between two arms. If she chooses arm a then she generates a payoff $d\pi_i(t) = w^a\,dt$ and a signal vector $d\lambda_i(t) = \sqrt{\xi^a}\mu\,dt + dZ_i(t)$. These choices are made simultaneously and independently. All players then observe all choices and all signals. No payoffs are observed. Here: $a \in \{1,2\}$; $d\pi_i(t) \in R$; $w^a \in \{w_1^a, w_2^a, ..., w_L^a\} \subset R$; L is the number of states of the world; $d\lambda_i(t) \in R^K$; K is the number of signals; $\mu \in \{\mu_1, \mu_2, ..., \mu_L\} \subset R^K$; $\xi^a > 0$ is the quality of the signal vector associated with arm a; the $dZ_i(t)$ are independently and identically distributed with distribution $\mathrm{N}[0, I_K\,dt]$ for $0 \leq i \leq N$; and I_K denotes the K by K identity matrix. Player i's objective is to maximize the expectation of the present discounted value of his payoff stream, namely

$$\mathrm{E}\left[\int_0^\infty re^{-rt}d\pi_i(t)\right],$$

where $r > 0$ is the discount rate.

Remark 7 *A background signal vector $d\lambda_0(t) = \sqrt{\xi^0}\mu\,dt + dZ_0(t)$ with $\xi^0 \geq 0$ could also be included in the model. This does not, however, involve any real gain in generality. For example, one could simply add further signals to each arm in an equivalent way.*

Remark 8 *There is no loss of generality in assuming that the components of $dZ_i(t)$ are independent and identically distributed. Indeed, suppose that $dZ_i(t) \sim \mathrm{N}[0, V\,dt]$, where V is a positive definite K by K matrix. Then $d\lambda_i(t)$ is observationally equivalent to*

$$V^{-1/2}\,d\lambda_i(t) = \sqrt{\xi^a}\,V^{-1/2}\mu\,dt + V^{-1/2}\,dZ_i(t),$$

and $V^{-1/2}\,dZ_i(t) \sim \mathrm{N}[0, I_K\,dt]$. What is important is that the distribution of $dZ_i(t)$ be non-degenerate. If it is not, then the players may be able to obtain exact – as opposed to noisy – information about the state of the world by looking at an appropriate linear combination of the signals.

13.2 The Filtering Argument

Suppose that the posterior concerning the state of the world l is given by $p \in R^L$, with $\sum_{l=1}^L p_l = 1$. Suppose furthermore that player i chooses $x_i \in \{0,1\}$,

with $x_i = 0$ if arm 1 is chosen, and $x_i = 1$ if arm 2 is chosen. Then it can be shown that

$$dp \sim \mathrm{N}\left[0, \left(\left(N - \sum_{i=1}^{N} x_i\right)\xi^1 + \left(\sum_{i=1}^{N} x_i\right)\xi^2\right)\Sigma(p)\,dt\right],$$

where

$$\Sigma_{l_1 l_2}(p) = \sum_{k=1}^{K} p_{l_1} p_{l_2} \left(\mu_{l_1 k} - \mu_{pk}\right)\left(\mu_{l_2 k} - \mu_{pk}\right)$$

and $\mu_p = \sum_{l=1}^{L} p_l \mu_l$. C.f. Theorem 9.1 of Liptser and Shiryayev [16]. Moreover, the expectation of player i's current payoff is

$$\Big((1 - x_i)\, w^1(p) + x_i w^2(p)\Big)\, dt,$$

where $w^a(p) = \sum_{l=1}^{L} p_l w_l^a$. Hence the original problem is equivalent to a new problem in which the players control the variance of the stochastic process p in such a way as to maximize their respective payoffs

$$\mathrm{E}\left[\int_0^\infty r e^{-rt}\Big((1 - x_i(t))\, w^1(p(t)) + x_i(t) w^2(p(t))\Big)\, dt\right],$$

where $x_i(t)$ is the action chosen by player i at time t. (Note that $x_i(t)$ also depends on the history of the process p.)

Remark 9 *It is now easy to see that the original model is effectively a special case of the present model. One need only set* $K = 1$, $L = 2$, $\xi^1 = \frac{\xi}{N}$, $\xi^2 = 1 + \frac{\xi}{N}$, $w_1^1 = w_2^1 = s$, $w_1^2 = \ell$, $w_2^2 = h$, $\mu_{11} = \ell/\sigma$ *and* $\mu_{21} = h/\sigma$.

13.3 The Time-Change Argument

We can transform the controlled-variance problem to a randomized-stopping problem by introducing a new time variable $\widetilde{t}$ according to the formula

$$d\widetilde{t} = \left(\left(N - \sum_{i=1}^{N} x_i\right)\xi^1 + \left(\sum_{i=1}^{N} x_i\right)\xi^2\right) dt.$$

In the new problem:

- there is no discounting;

- there are no flow payoffs;
- at each instant of time, the change in the posterior

$$dp \sim \mathrm{N}\left[0, \Sigma(p)\, d\widetilde{t}\right];$$

- at each instant of time, the game comes to an end with probability

$$\frac{r\, d\widetilde{t}}{\left(N - \sum_{i=1}^{N} x_i\right)\xi^1 + \left(\sum_{i=1}^{N} x_i\right)\xi^2};$$

- if the game does come to an end, player i receives a lump-sum payoff of

$$(1 - x_i)\, w^1(p) + x_i w^2(p).$$

Remark 10 *The randomized-stopping problem introduced in this section and the randomized-stopping problem used in Section 5 are not identical, but they are closely related. In fact, it is easy to check that the Bellman equation for the team problem considered in Section 5 can be written in the form*

$$0 = \max_{n \in \{0,1,\ldots,N\}} \left(\frac{r}{\xi + n} \left(\left(1 - \frac{n}{N}\right)(s - u_*) + \frac{n}{N}(w - u_*) \right) + \Sigma \frac{(u_* - s)''}{2} \right),$$

which corresponds to a randomized-stopping problem of the type introduced in the present section. Similarly, the team version of the randomized-stopping problem considered in the present section is equivalent to a randomized-stopping problem in which: (i) all payoffs are measured relative to the payoff w^1; (ii) there is no discounting; (iii) the state variable p evolves according to the equation $dp \sim \mathrm{N}\left[0, \Sigma\left(p\right) d\widetilde{t}\right]$ until such time as the process is stopped; (iv) the flow payoff $-\frac{r\left(w^1 - w^2\right)}{N(\xi^2 - \xi^1)}$ is obtained up to the point at which the process is stopped; (v) the process is stopped with probability $\frac{r}{(N-n)\xi^1 + n\xi^2}$ per unit time; (vi) when the process is stopped, a lump-sum payoff of $\frac{\xi^1\left(w^1 - w^2\right)}{\xi^2 - \xi^1}$ is obtained. The advantage of the formulation used in Section 5 is that it leads quickly to a characterization of the optimal policy. The advantage of the formulation used in the present section is that it treats the two arms symmetrically, and therefore leads to a randomized-stopping problem, the relevance of which is easier to understand. (We did not introduce both problems in Section 5 because we did not want to overburden the exposition.)

13.4 The Team Problem

We make the following assumptions:

(i) $\xi^2 > \xi^1$;

(ii) there exist l_1 and l_2 such that $w^1_{l_1} > w^2_{l_1}$ and $w^2_{l_2} > w^1_{l_2}$;

(iii) $\mu_{l_1} = \mu_{l_2}$ iff $l_1 = l_2$.

In other words, arm 2 is more informative than arm 1, neither arm dominates the other in payoff terms, and no two states of the world are indistinguishable.

These assumptions are not substantive: if both arms are equally informative, then the team will choose the arm with the highest short-run payoff; if one arm dominates the other in payoff terms, then that arm will always be chosen; and if two states of the world are indistinguishable, then they can be combined into a single composite state, thereby reducing the dimensionality of the problem by one.

Let u_* be the value function for the team problem, let $u_\infty(p) = \max\{w^1(p), w^2(p)\}$ be the myopic payoff, and let $u_0(p) = \sum_{l=1}^{L} p_l \max\{w^1_l, w^2_l\}$ be the full-information payoff. Then: u_* is convex in p; u_* is Lipschitz continuous jointly in p, r, ξ^1, ξ^2 and N; $u_\infty \leq u_* \leq u_0$ (with strict inequality for all interior p); u_* is non-increasing in r (strictly decreasing for all interior p), $u_* \to u_0$ as $r \to 0$, and $u_* \to u_\infty$ as $r \to \infty$; u_* is non-decreasing in N (strictly increasing for all interior p), and $u_* \to u_0$ as $N \to \infty$; u_* is non-decreasing in ξ^1 and ξ^2 (strictly increasing for all interior p), and $u_* \to u_0$ as $\xi^2 \to \infty$.

As for the amount of experimentation, it is a dominant strategy to play arm 2 when $w^2 > w^1$. It therefore makes sense to say that experimentation occurs if arm 2 is played when $w^1 > w^2$. Now it follows from the randomized-stopping formulation of the team problem that the team chooses arm 2 iff

$$\frac{w^2 - u_*}{\xi^2} \geq \frac{w^1 - u_*}{\xi^1}$$

or

$$u_* \geq \frac{\xi^2 w^1 - \xi^1 w^2}{\xi^2 - \xi^1}.$$

Hence the experimentation region: is convex; is strictly decreasing in r, converges to the zero-discounting experimentation region as $r \to 0$, and vanishes

as $r \to \infty$; is strictly increasing in N, and converges to the entire region $\{w^1 > w^2\}$ as $N \to \infty$; is strictly increasing in ξ^2, and converges to the entire region $\{w^1 > w^2\}$ as $\xi^2 \to \infty$.

Remark 11 *Each of the three different formulations of the team problem has advantages. For example, that u_* is convex follows at once from the original optimal-experimentation formulation of the team problem, and that an optimal strategy exists follows most easily from the randomized-stopping formulation. The randomized-stopping formulation is, however, probably the most useful. For example, this formulation allows one to obtain explicit formulae for the rate of change of u_* with r, N, ξ^1 and ξ^2.*

Remark 12 *Somewhat stronger results are available in the case $L = 2$. In this case: u_* is twice continuously differentiable in p; u_* is strictly convex in the sense that $u_*'' > 0$; and the experimentation region is strictly decreasing in ξ^1.*

Remark 13 *It is difficult to determine the variation of the experimentation region with ξ^1 in the general case. Increases in ξ^1 tend to increase the experimentation region via an encouragement effect, and to decrease the experimentation region via a free-rider effect. However, by contrast with the case $L = 2$, it is difficult to determine which of these effects dominates. Indeed, as far as we know the experimentation region may grow in some places and shrink in others.*

Remark 14 *The proof that u_* is Lipschitz continuous jointly in p, r, ξ^1, ξ^2 and N follows the lines of Krylov [15], Sections 3.1 and 4.1.*

13.5 The Strategic Problem

In order to ensure the existence of a symmetric equilibrium, it is necessary to allow for mixing. That is, we must replace a player's choice of an action $x_i \in \{0, 1\}$ by the choice of a probability $X_i \in [0, 1]$. Doing so leaves all the formulae given so far in the present section unchanged, except that the symbol 'x' must be replaced by the symbol 'X'.

The strategic problem exhibits the same pair of complementarities as before: increasing the total experimentation of the other players increases a

player's payoff, and increasing her payoff increases the amount of experimentation that she is willing to undertake. By exploiting these, one can show that it possesses a maximal symmetric equilibrium. Suppose that $u_\#$ is the value function of this equilibrium. Then: $u_\infty \leq u_\# \leq u_*$ (with strict inequality for all interior p); $u_\#$ is non-increasing in r (strictly decreasing for all interior p), $u_\# \to u_0$ as $r \to 0$, and $u_\# \to u_\infty$ as $r \to \infty$; $u_\#$ is non-decreasing in N (strictly increasing for all interior p), and $u_\#$ converges to a limit less than u_0 as $N \to \infty$ (strictly less than u_0 for all interior p); $u_\#$ is non-decreasing in both ξ^1 and ξ^2 (strictly increasing for all interior p), and $u_\# \to u_0$ as $\xi^2 \to \infty$. Similarly, one can show that the experimentation region: is strictly decreasing in r, converges to the zero-discounting experimentation region as $r \to 0$, and vanishes as $r \to \infty$; is strictly increasing in N, and converges to a limiting region strictly smaller than the region $\{w^1 > w^2\}$ as $N \to \infty$; and is strictly increasing in ξ^2, and converges to the entire region $\{w^1 > w^2\}$ as $\xi^2 \to \infty$.

Remark 15 *Stronger results are available in the case $L = 2$. In this case $u_\#$ is twice continuously differentiable in p, and $u_\#$ is strictly convex in the sense that $u''_\# > 0$. In the general case, all we can show is that $u_\#$ is measurable, and there does not appear to be any reason why $u_\#$ should be convex.*

13.6 The Non-Degenerate Case

Associated with each signal k there is a contribution $\Sigma^k(p)\,dt$ to the covariance matrix of the posterior, where

$$\Sigma^k_{l_1 l_2}(p) = p_{l_1} p_{l_2} \left(\mu_{l_1 k} - \mu_{pk}\right)\left(\mu_{l_2 k} - \mu_{pk}\right).$$

In other words, signal k causes the posterior to move in the direction $\pm\sigma^k(p)$, where

$$\sigma^k_l(p) = p_l \left(\mu_{lk} - \mu_{pk}\right).$$

Hence, if rank $\left\{\sigma^1(p), \sigma^2(p), ..., \sigma^K(p)\right\} = L-1$, then p can move in all possible directions within the unit simplex in R^L. Let M be the L by $K+1$ matrix with entries $M_{lk} = \mu_{lk}$ for $1 \leq l \leq L$ and $1 \leq k \leq K$, and $M_{(K+1)l} = 1$ for $1 \leq l \leq L$. Then it is easy to check that rank $\left\{\sigma^1(p), \sigma^2(p), ..., \sigma^K(p)\right\} = L-1$ for all interior p iff rank $(M) = L$. We say that the game is *non-degenerate* if this latter condition holds.

When the game is non-degenerate, the value function for the team problem is twice continuously differentiable, and is the unique bounded solution of the Bellman equation

$$0 = \max_{n\in\{0,1,...,N\}} \left(\frac{r}{(N-n)\,\xi^1 + n\xi^2} \right.$$
$$\times \left(\left(1 - \frac{n}{N}\right) \left(w^1(p) - u_*(p)\right) + \frac{n}{N} \left(w^2(p) - u_*(p)\right) \right)$$
$$\left. + \frac{1}{2} \sum_{l_1,l_2=1}^{L} \Sigma_{l_1 l_2}(p) \frac{\partial^2 u_*}{\partial p_{l_1} \partial p_{l_2}}(p) \right).$$

Similarly, the value function $u_{\#}$ of the maximal symmetric equilibrium is twice continuously differentiable, and is a bounded solution of the Bellman equation

$$0 = \frac{r}{(N - T(p, u_\dagger(p)))\,\xi^1 + T(p, u_\dagger(p))\,\xi^2}$$
$$\times \left(\left(1 - \frac{T(p, u_\dagger(p))}{N}\right) \left(w^1(p) - u_\dagger(p)\right) + \frac{T(p, u_\dagger(p))}{N} \left(w^2(p) - u_\dagger(p)\right) \right)$$
$$+ \frac{1}{2} \sum_{l_1,l_2=1}^{L} \Sigma_{l_1 l_2}(p) \frac{\partial^2 u_\dagger}{\partial p_{l_1} \partial p_{l_2}}(p),$$

where

$$T(p,u) = \left\{ \begin{array}{ccc} 0 & \text{if} & A(p,u) < 0 \text{ and } w^1(p) > w^2(p) \\ \dfrac{NA(p,u)}{N-1} & \text{if} & A(p,u) \in [0, N-1] \text{ and } w^1(p) > w^2(p) \\ N & \text{if} & A(p,u) > N-1 \text{ or } w^1(p) \leq w^2(p) \end{array} \right\}$$

and

$$A(p,u) = \frac{u - w^1(p)}{w^1(p) - w^2(p)} - \frac{N\xi^1}{\xi^2 - \xi^1}.$$

Remark 16 *The assumption that $\mu_1 \neq \mu_2$ automatically implies that the game is non-degenerate when $L = 2$. This explains the smoothness results described for this case in Sections 13.4 and 13.5.*

14 Conclusion

In this paper we have analyzed team and equilibrium experimentation in a many-player common-value two-armed bandit problem in terms of the free-rider effect and the encouragement effect. Our analysis allowed for a very general specification of the number of states of the world, and of the pattern of information generated by the two arms. It did not, however, allow for many arms, nor did it allow the pattern of information generated by one arm to differ from the pattern of information generated by the other. Allowing for many arms, all yielding the same pattern of information, raises one new issue: what is the best incentive-compatible pattern of experimentation? Continuing to restrict attention to two arms, but allowing the pattern of information generated by one arm to differ from the pattern of information generated by the other, raises another: what is the best way of exploiting the different patterns of information generated by the two arms? We hope to address these issues in future work.

References

[1] Aghion, P., Bolton, P., Harris, C., and Jullien, B. (1991), "Optimal Learning by Experimentation", *Review of Economic Studies*, 58, 621-654

[2] Aghion, P., Paz-Espinoza, M., and Jullien, B. (1993), "Dynamic Duopoly with Learning through Market Experimentation", *Economic Theory*, 3, 517-539

[3] Banerjee, A. V. (1992), "A Simple Model of Herd Behaviour", *Quarterly Journal of Economics*, CVII, 797-817

[4] Berry, D. A., and Fristedt, B. (1985), "Bandit Problems", Chapman and Hall, New York

[5] Bhattacharya, S., Chatterjee, K., and Samuelson, L. (1986), "Sequential Research and the Adoption of Innovations", *Oxford Economic Papers*, Special Issue, 38, 218-243

[6] Bikhchandani, S., Hirshleifer, D., and Welch, I. (1992), "A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades", *Journal of Political Economy*, 100, Vol. 5, 992-1026

[7] Bolton, P. and Harris, C. (1993), "Strategic Experimentation", STICERD Discussion paper TE/93/261, London School of Economics

[8] Chamley, C., and Gale, D. (1994), "Information Revelation and Strategic Delay in a Model of Investment", *Econometrica*, 62, 1065-1086

[9] Easley, D., and Kiefer, N. (1988), "Controlling a Stochastic Process with Unknown Parameters", *Econometrica*, 56, 1045-1064

[10] Ellison, G., and Fudenberg, D. (1993), "Rules of Thumb for Social Learning", *Journal of Political Economy*, 101, 612-644

[11] Freidlin, M. (1985), "Functional Integration and Partial Differential Equations", Annals of Mathematics Studies 109, Princeton University Press.

[12] Harris, C. (1993), "Generalized Solutions of Stochastic Differential Games in One Dimension", Industry Studies Program Discussion Paper # 44, Boston University, December

[13] Hendricks, K., and D. Kovenock (1989), "Asymmetric Information, Information Externalities, and Efficiency: the Case of Oil Exploration", *Rand Journal of Economics*, 20, No2, 164-182

[14] Karatzas, I. (1984), "Gittins Indices in the Dynamic Allocation Problem for Diffusion Processes", *Annals of Probability*, 12, 173-192

[15] Krylov, N.V. (1980), "Controlled Diffusion Processes", Springer Verlag, New York, Heidelberg, Berlin

[16] Liptser, R.S., and Shiryayev, A.N. (1977), "Statistics of Random Processes I", Springer Verlag, New York, Heidelberg, Berlin

[17] McLennan, A. (1984), "Price Dispersion and Incomplete Learning in the Long Run", *Journal of Economic dynamics and control*, 7, 331-347

[18] Mirman, L., Samuelson, L., and Urbano, A. (1989), "Duopoly Signal Jamming", University of Virginia Working Paper, Department of Economics

[19] Rob, R. (1991), "Learning and Capacity Expansion under Demand Uncertainty", *Review of Economic Studies*, 58, 655-675

[20] Rothschild, M. (1974), "A Two-Armed Bandit Theory of Market Pricing", *Journal of Economic Theory*, 9, 185-202

[21] Shiryayev, A.N. (1978), "Optimal Stopping Rules", Springer Verlag, New York, Heidelberg, Berlin

[22] Smith, L. (1991), "Error Persistence, and Experiential versus Observational Learning", Foerder Discussion Paper, Tel-Aviv University

[23] Vives, X. (1992), "Learning from Others", mimeo, Institut d'Anàlisi Econòmica Universitat Autònoma de Barcelona, December